DOCUMENT RESUME

ED 206 693 TM 810 592

AUTHOR	Caulley, Darrel N., Ed.
TITLE	The Relationships Between Policy and Evaluation in State Departments of Education. Research on Evaluation Program Paper and Report Series.
INSTITUTION	Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY	National Inst. of Education (ED), Washington, D.C.
REPORT NO	NWREL-REP-52
PUB DATE	Jan 81
CONTRACT	400-80-0105
NOTE	198p.
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	*Evaluation; Evaluation Methods; *Policy Formation; Public Policy; *State Departments of Education
IDENTIFIERS	Michigan; Oregon; *Policy Analysis; South Carolina; Virginia; Washington; Wisconsin

ABSTRACT

What are the relationships between policy and evaluation within state departments of education? How does policy influence what is evaluated in such settings and how does evaluation inform policymaking? Directors of evaluation in six state education agencies address these issues through detailed descriptions of their operations: Gordon Ascher, Oregon; Gerald Bracey, Virginia; David L. Donovan, and Stanley A. Rumbaugh, Michigan; James H. Gold, Wisconsin; Alfred Rasp, Jr., Washington; and Paul D. Sandifer, South Carolina. An educational philosopher specializing in policy analysis, Thomas F. Green, then provides a commentary on the preceding six papers. (Author/BW)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED206693

TM 820 612

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

paper and report series

Research on Evaluation Program

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
M. M. Rogers
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204
Telephone: (503) 248-6800

No. 52 THE RELATIONSHIPS BETWEEN POLICY AND EVALUATION IN STATE DEPARTMENTS OF EDUCATION

DARREL N. CAULLEY, Editor

January 1981

Nick L. Smith, Director
Research on Evaluation Program
Northwest Regional Educational Laboratory
710 S.W. Second Avenue, Portland, Oregon 97204

Published by the Northwest Regional Educational Laboratory, a private nonprofit corporation. The work upon which this publication is based was performed pursuant to Contract No. 400-80-0105 of the National Institute of Education. It does not, however, necessarily reflect the views of that agency.

PREFACE

The Research on Evaluation Program is a Northwest Regional Educational Laboratory project of research, development, testing, and training designed to create new evaluation methodologies for use in education. This document is one of a series of papers and reports produced by program staff, visiting scholars, adjunct scholars, and project collaborators--all members of a cooperative network of colleagues working on the development of new methodologies.

What are the relationships between policy and evaluation within state departments of education? How does policy influence what is evaluated in such settings and how does evaluation inform policymaking? Directors of evaluation in six state education agencies (SEAs) address these issues through detailed descriptions of their operations. An educational philosopher specializing in policy analysis then provides a commentary on the preceding six papers.

Nick L. Smith, Editor
Paper and Report Series

CONTENTS

Page

Introduction 1

The Interaction Between Policy and Evaluation Within Oregon Department of Education (Gordon Ascher) 3

The Relationship Between Evaluation and Policy In Educational Decisions in Virginia (Gerald Bracey) 28

The Interaction Between Policy and Evaluation Within the Michigan Department of Education (David L. Donovan and Stanley A. Rumbaugh) 55

The Interaction Between Policy and Evaluation in State Departments of Education (James H. Gold) 82

Evaluation and Decision Making in Washington State (Alfred Rasp, Jr.) 113

The Interaction Between Policy and Evaluation Within the South Carolina Department of Education (Paul D. Sandifer) 142

Policy and Evaluation: A Conceptual Study (Thomas F. Green) 162

THE RELATIONSHIPS BETWEEN POLICY AND EVALUATION IN STATE DEPARTMENTS OF EDUCATION

Introduction

With the increased interest in improving the utilization of evaluation results, some researchers have begun to more closely examine the relationships between policy and evaluation. This topic is of particular concern to the staff of the Research on Evaluation Program because of our interest in developing new methods which increase the utilization of evaluation. We therefore sought to understand better how policy influences evaluation and how evaluation in turn shapes policy. Instead of examining these issues through somewhat artificial surveys or field interviews, we decided to have the most knowledgeable individuals, evaluation managers themselves, discuss their own insights concerning these matters.

Evaluators in six state departments of education were commissioned to prepare reports on the interface between policy and evaluation in their settings. This document thus contains the reports of the following individuals.

Gordon Ascher	Oregon State Department of Education
Gerald Bracey	Virginia State Department of Education
David Donovan and Stanley Rumbaugh	Michigan State Department of Education
James Gold	Wisconsin State Department of Education
Alfred Rasp	Washington State Department of Education
Paul Sandifer	South Carolina State Department of Education

To orient the writing of those authors, the following questions were suggested as the focus of their reports:

- How does policy influence evaluation? For example: How does policy determine what is evaluated? Does policy somehow influence evaluation methods? How does policy influence the nature and organization of your evaluation unit? To what extent do policymakers call on evaluators to help them with their policy problems? What policy problems require the use of evaluation information? What percentage of evaluators' time is devoted to policy questions?

- How does evaluation influence policy? For example: To what extent is evaluation used in policymaking? How is evaluative information communicated to policymakers? How is evaluative information used in arriving at policy decisions?

Thomas F. Green, an educational philosopher from Syracuse University was commissioned to provide a general discussion of policy formation and to comment on the six reports. After the authors had the opportunity to review each other's work at an invitational conference, final drafts of the papers were prepared for inclusion here.

Darrel N. Caulley, who assembled these materials, originally intended to supplement the collection with an additional analysis and commentary designed to highlight the common themes of the six papers. Due to a serious and lengthy illness, however, he has been unable to do so. Hopefully, he will be able to return to that task in a few months.

These papers present several insiders' views of the nature of evaluation practice in state departments of education. Their rich examples help illuminate the difficulties of providing useful evaluations in state agency settings. Readers interested in further discussion of these papers are referred to ROEP Report #46, "The Context of Evaluation Practice in State Departments of Education" by Nick L. Smith. This latter report draws on these six papers to describe the forces that shape the practice of evaluation in state departments. This latter paper thus contains additional analyses of the reports contained herein.

Nick L. Smith

8

THE INTERACTION BETWEEN POLICY AND EVALUATION
WITHIN OREGON DEPARTMENT OF EDUCATION

Gordon Ascher
Associate State Superintendent of Public Instruction
Oregon State Department of Education

A Conceptual Framework

Policy and Evaluation Defined

It is not the purpose of this paper to develop a complete conceptual framework for policy and evaluation. Its purpose, rather, is to show, in a practical setting, how these two interact. But it is recognized that some definition is necessary in order to provide a common understanding which will permit communication between author and reader. For the purpose of this paper, then, the following definitions are presented:

Policy: Guidance provided by an organization to decision makers.

Evaluation: The collection of information for use in making decisions.

The common factor in both definitions is "decision making." In some manner, both policy and evaluation are used by a person who is forced to make a decision. How they interact will, of course, vary with the situation. But the fact that they do interact is a major premise here and the fact that their interaction occurs in the decision making process is a second major premise. Therefore, we can only meaningfully define these terms dynamically; by considering how they function jointly and severally in their mutual environment of decision making.

The Interaction of Policy and Evaluation

Policy and evaluation both serve the decision making process. They interact with each other in a variety of ways when decisions are made. To illustrate, consider an unhappy consumer who finds that the radio he recently purchased does not work. He returns to the store seeking to return the radio. What possible reactions can he get from the salesperson? The salesperson will either take the radio back or not. How the decision is made depends on how the store uses policy and evaluation.

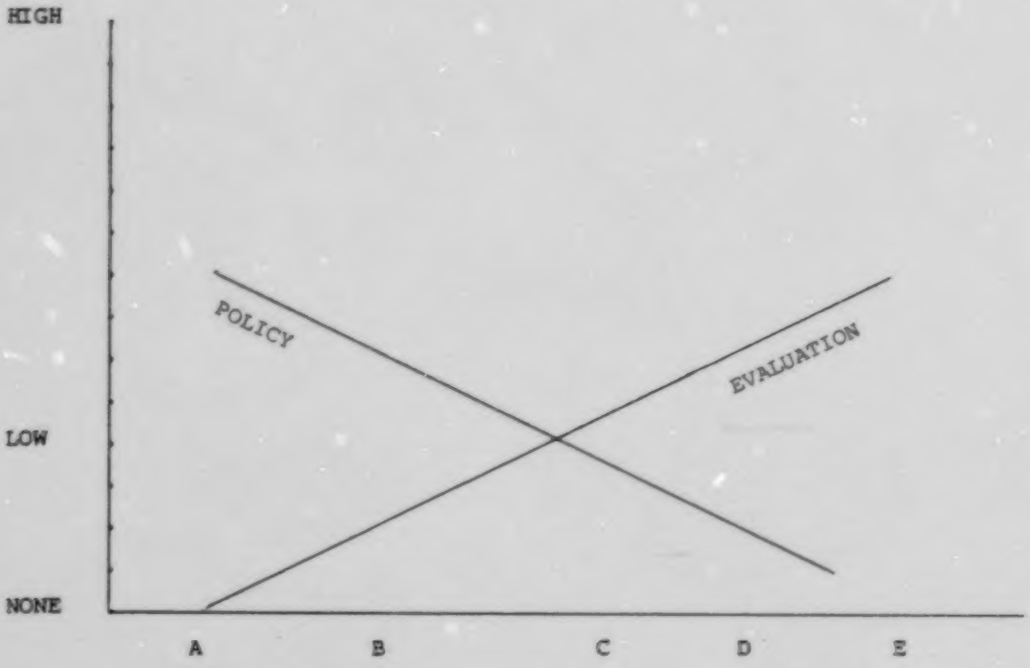
Scenario 1: The salesperson says: "It is our policy to accept all (or no) returns, no questions asked." Here the decision appears to be based wholly on policy. An evaluation of how the radio functions is not required to make the decision. We should, however, suspect that evaluation did have some influence on the adoption of a policy to accept all returns. Probably, market research or the storeowner's own informal collection of information indicated that such a policy is, in the long run, good business.

Scenario 2: The salesperson says: "It is our policy to decide all returns based on our evaluation of the product. We will have our service department conduct the evaluation, and then we will decide." The decision is based on policy and on information generated through product evaluation. That is, a blend of evaluation and policy.

Scenario 3: The salesperson says: "We don't have a general policy on returns, but the radio is obviously defective, so you can return it." Having evaluated the radio's function, it appears that the salesperson is able to make a decision based only on evaluation and with no policy influence. But the absence of policy is illusory. Implicit in the salesperson's willingness to take back a defective product is a policy to be fair, or to please the customer, or to avoid legal action by recognizing an implied warranty.

It appears from the scenarios that no decision is made without the influence of both policy and evaluation. Why, then, do our real life experiences lead us to think that some decisions can be made without the influence of policy; or without the influence of evaluation? The answer lies in our perception of the influences of policy and evaluation relative to each other on the making of the decision. Our perception of the relative influence of each is determined by the proximity of the influence to the decision situation. A graphic presentation of this model of the perceived relative influences of the making of a decision appears in Figure 1:

Figure 1



At Point B, a decision maker would perceive that her decision is influenced to a greater degree by policy than by evaluation. Similarly, at Point D she would perceive being influenced to a greater degree by evaluation than by policy. At Point C, the perceived influences are at equipoise. Points A and E do not exist in reality because every decision is influenced by both policy and evaluation. The misleading perceived relative influences are determined by the proximity of the influence to the immediate decision to be made.

Point B represents Scenario 1. The perceived influence of evaluation by the salesman is nonexistent. Actually, the influence of evaluation exists but is perceived to be so small relative to the influence of policy, it appears nonexistent when the salesman says: "It is our policy to accept all returns." The perception of the influence is determined by the proximity of the influence to the decision situation. The reason it appears to the salesman that evaluation is not necessary for his decision to take the radio back, is that he is thinking in terms of evaluation of the radio, the object to be returned. He does not perceive the influence of a more remote evaluation: The store owner's market research which led him to believe that the general policy of accepting all returns is good business. So evaluation influenced the policy directly, but influenced the immediate decision about the radio only remotely and so it appeared to the decision maker that evaluation did not influence his decision. Thus, it often appears to us that policy (or evaluation) does not enter into the decisions we make. However, if we look hartd enough we find the remote influence of evaluation (or policy) in every decision we make. The model, just described, can be summarized by ten "rules":

1. The interaction of policy and evaluation occurs during the decision making process.

2. The nature of this interaction is determined by the relative influence each has on the making of the decision.

3. The Perceived influence of policy and evaluation on the making of a decision is determined by the proximity of the influence to the decision making situation.

4. A decision can be made with the direct influence of both policy and evaluation in the decision situation (Scenario 2).

5. A decision can be made with the direct influence of policy in the decision situation and with no direct influence of evaluation. The influence of evaluation on the decision exists but is remote to the decision situation. A total lack of evaluation influence is illusory (Scenario 1).

6. Similarly, a decision can be made with the direct influence of evaluation in the decision situation and with no direct influence of policy. The influence of

policy on the decision exists but is remote to the decision situation. A total lack of policy influence is illusory (Scenario 3).

7. Policy is created or affirmed each time a decision is made. It is common when making a decision to look at earlier decisions made under a similar set of circumstances. If the later decision is the same as the earlier decision, the policy which influenced the earlier decision is affirmed. If the later decision rejected the influence of the policy applied in the earlier decision, new policy may be gleaned from the new decision. Court decisions function in this way.*
8. Evaluation information is unaffected by a decision it influenced. The weight and credibility of the data in future use, however, may be affected by how the decision maker allowed it to influence the decision.
9. Policy has no influence on decision making unless the policy is communicated and known to the decision maker.
10. Similarly, evaluation information has no influence on a decision unless the information is communicated and is known to (and understood by) the decision maker.

Policy and Evaluation in the Oregon Department of Education

The ten rules present the essence of the policy/evaluation interaction model. The remainder of the paper presents examples of this interaction in the Oregon Department of Education. For these examples to be understood, it is first necessary to present some description of the Department organization and how it functions.

---

*See, e.g., Caulley, D. and Dowdy, I. Evaluation case histories as a parallel to legal case histories. Portland, Oregon: Northwest Regional Educational Laboratory.

Organization of the Oregon Department of Education

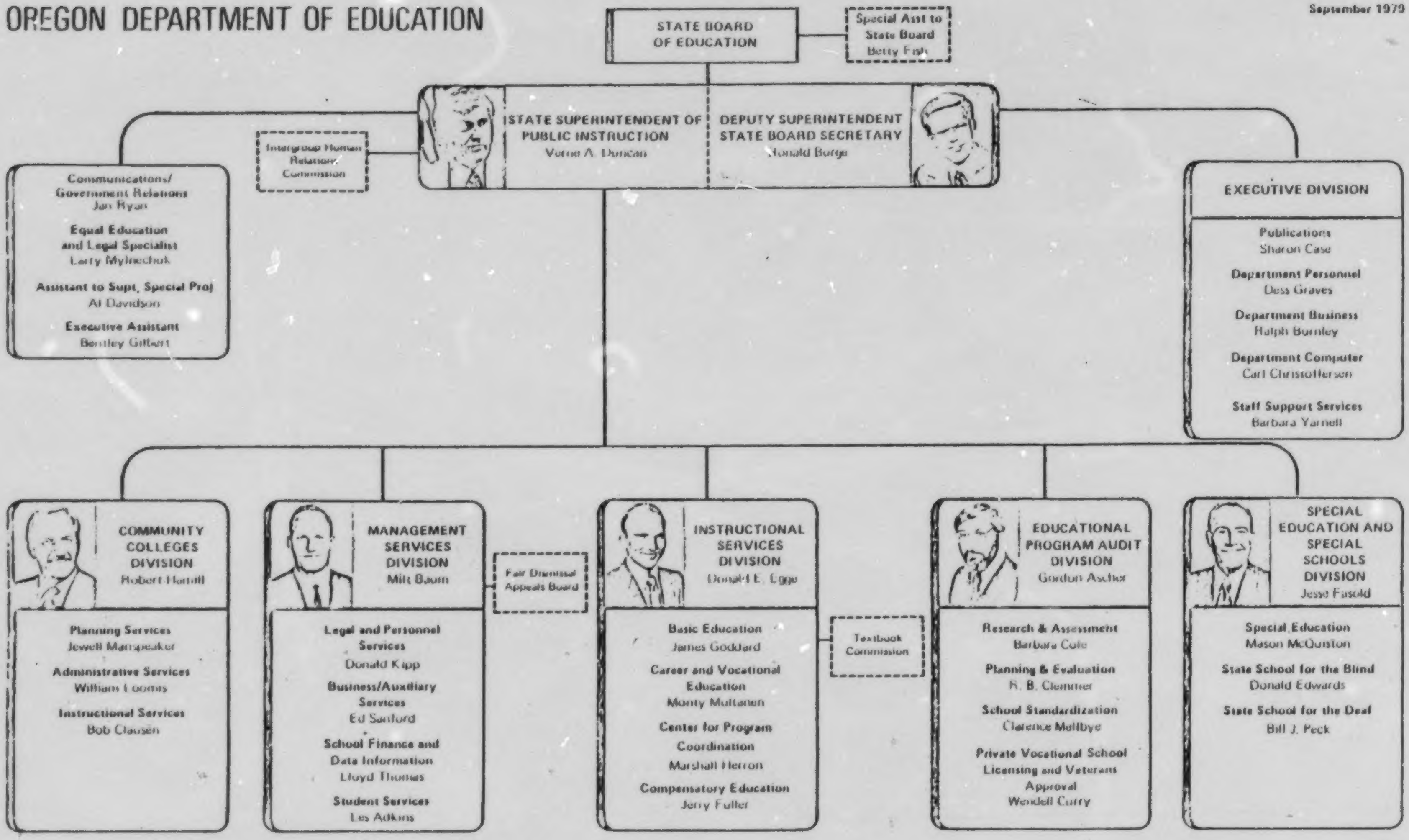
Figure 2 presents the Department's table of organization. Note that the five associate state superintendents report directly to the State Superintendent of Public Instruction and not to the Deputy State Superintendent. This has an effect on the making of policy. Further, you will note that both the Superintendent and the Deputy, like each associate, manages a division. Thus, they are involved in day-to-day operations and do not exist in an ivory tower. This has an influence on policy making. The Educational Program Audit Division was established by Superintendent Duncan to separate evaluation responsibilities from program responsibilities. This was to relieve the ambivalence of program support people (e.g. Title I field liaison people) who put a great deal of energy into helping local district personnel develop programs to aid children and later have to apply the (often antithetical) evaluation rules developed by federal agencies or others in the Department. For example, this Division monitors P.L. 94-142 compliance while the Special Education Division provides technical support for program development in the field. Additional examples are possible but the point is made. This innovation has worked to the satisfaction of all parties and has had a great deal of influence on thinking about and doing evaluation-based decision-making throughout the Department. While the remaining divisions do not have the evaluation responsibility of the Program Audit Division. Yet each still must do a great deal of evaluation within its program support functions and in this case the Program Audit Division assists by providing technical assistance.

The State Board of Education, a lay board appointed by the Governor with the approval of the Senate, has been assigned a policy making function by the legislature. The State Superintendent, an elected official, carries out the Board's policies. In doing so he creates a great deal of policy by interpreting Board policy and in order to fill gaps in policy on issues not addressed by the Board. The Superintendent's powers are unusual in that he performs the same function as a district

FIGURE 2

OREGON DEPARTMENT OF EDUCATION

September 1979



or circuit court in interpreting state and federal statutes concerning education. Appeal from his decision is made directly to the appellate division of the courts.

The policies of the Board reside in the administrative rules promulgaged by the Board. These follow a legislatively prescribed procedure and have the same weight as legislatively enacted statutes. Board policies in many areas have not been delineated, but must be gleaned from the administrative rules. While the Board is currently considering a process to develop a set of policy statements, Oregon practicality dictates that effort produce usable results. Flowery statements, or broad generalizations presented by other states do little to provide the guidance to decision makers that the Oregon Minimum Standards for Elementary and Secondary Schools (and the other sets of standards for other schools) provide. The policies of the Superintendent are found in administrative memoranda and other special memoranda he promulgates and policy may be gleaned from his decisions in controversies (a body of quasi-case law).

The Policy Makers

While it is correct to say that the State Board of Education and the State Superintendent are the policy makers identified by the legislature, the reality is that many others make policy when carrying out functions delegated to them by the Board and the Superintendent. Therefore, to look at policy in the Department, we must look beyond the table of organization for a functional description.

Superintendent Duncan has established a cabinet. It consists of the Deputy Superintendent and all of the associate superintendents. This cabinet meets several times each month as the Superintendent's Council. Note that the apostrophe in "Superintendent's" is placed before the "s" and not following the "s." Thus, the Council is advisory to the Superintendent. Superintendent Duncan maintains his freedom to hear the Council's advice and disregard it if a higher wisdom so dictates. The council, however, is not at all "window dressing." The discussions and the arguments put forth are a serious part of the decision making process and also serve to keep the associates current as to the Superintendent's policy. This policy provides guidance to the associates when they make the many decisions required of them in the operation of their divisions. The council also serves to inform the Superintendent of decisions made by associates (and other staff). These decisions have established policy and, since the associates acted as the Superintendent's agents, the decisions somewhat limit the Superintendent's freedom to establish policy. This balance of policy making roles is possible because of frequent formal and informal interaction among all parties involved.

Others who influence policy are the special interest groups. The Oregon Education Association, the Oregon School Boards Association and the Council of Oregon School Administrators are just three of many. The opinions of special interest groups are often solicited because they offer a perspective which Department staff may not have considered. The same is true of the opinions of citizens in general.

The legislature influences policy by enacting laws which dictate the limitations on policy in certain areas. More directly, the legislature influences policy through the budgetary process. The Department's budget is presented biennially to the legislature by the Board. The Board's policy to support programs for gifted students is given life by a dollar amount of support in the budget. The legislature, in modifying this allocation, places a limitation on the Board's application of its policy. One may argue that the policy to assist gifted students was not affected, only the degree of support was. But the next time the Board builds a budget, the earlier legislative action does affect policy.

The Evaluators

Those who hold payroll designations as "evaluators" reside in the Educational Program Audit Division. Approximately one percent of the total Department operational budget goes to support this division, but an estimated ten percent of the operational budget is used evaluation because every other division retains some evaluative function. So, just as policy making does not completely reside in an identifiable few, neither does evaluation. In fact, in reality, all policy makers evaluate or use evaluation information (formally or informally) and all who make decisions about the conduct of evaluation are influenced by policy and establish policy by making decisions about what to evaluate.

The Policy/Evaluation Interaction in the Oregon Department of Education

The Oregon Minimum Standards

Since decisions of all kinds on all levels of organization are made daily, many examples can be used to show how policy and evaluation interact in Oregon. Of all of these possibilities, one has been chosen for discussion on the basis of the fact that it involves everyone in the Department and the State Board of Education. The Elementary-Secondary Guide for Oregon Schools Part I (Oregon's Minimum Standards) has been chosen as an example of (1) a policy setting process which is influenced by evaluation information and (2) a statement of policy which influences evaluation. The Minimum Standards are discussed here because they are the Board's only complete statement about the Board's policy on evaluation and about the policy/evaluation interaction. Even though the Minimum Standards are applied at the local level, we use it here for illustration of the policy that is applied at the state agency level as well. Like many state agencies, there is not a complete policy statement for the workings of the agency itself. Informal transmittal of policy by the Board and the Superintendent indicate that the same policies apply at the state level, but there is not the concise form to show the reader.

The Minimum Standards are a set of administrative rules established by the State Board of Education under a legislative grant of authority to do so. These have the same weight and influence on the schools as do legislatively enacted statutes. The Board adopts many sets of administrative rtules but has chosen to designate only one subset the "Minimum Standards." While a school district is required to comply with all of the administrative rules, only failure to comply with a Minimum Standard will authomatically put into motion a process which requires the district to correct its deficiency or lose its state funding support. This process is the school standardization (accreditation) process administered by a section of the Educational Program Audit Division in response to a legislative mandate to determine that schools are meeting the standards set by the State Board. Of course, failure to comply with an administrative rule which is not part of the Minimum Standards subset carries penalties too. Those penalties, however, are stated in each rule or for a set of rules. But the school standardization process (team visits to every school in the state on a five-year cycle) is concerned only with compliance with the minimum standards.

Let us pause and see what we can detect so far concerning the policy/evaluation interaction.

— The Minimum Standards are the Board's expression of how a district must operate to provide a quality education to its students. These are statements of policy or rules based on policy.

— The legislative mandate that such policy be established was accompanied by a legislative mandate that the State Superintendent of Public Instruction determine whether schools are meeting these standards. This process of quality assurance is a process of evaluation.

— Thus, the legislature, desirous of quality schools, required clearly defined policy for schools to follow and evaluation to make sure they do. How does the legislature ensure that the policy and the evaluation will interact? By placing the Superintendent in a decision-making role. The decision he must make is which school districts will continue to receive the state funds provided by the legislature.

The discussion of the Minimum Standards now branches into two streams. The first stream is a consideration of what the Standards contain because that sheds light on how the Board is directing the school districts to carry out the Board's policy, and how evaluation is required to ensure that the policy is carried out. The second stream considers how the Board establishes these Standards, that is, how the Board determines policy and how evaluation information influences this policy making process. The discussion of the first stream, the contents of the Standards, is based on the Minimum Standards currently in use in Oregon and adopted by the Board on June 23, 1976. A revision process has been underway since 1978 and a new set of Standards will be adopted in 1980. However, for our purpose here, either set will do. The discussion of the second stream, how the Board establishes standards, is a discussion of the revision process which will result in the 1980 adoption.

The Role of Policy in the Conduct of Evaluation

The Content of the Minimum Standards

The 1976 Minimum Standards are divided into twelve sections:

-- Definitions. These are definitions of terms used in the Standards. They are more than definitions for the guidance of those involved with the standards because these have been adopted as an administrative rule much the same way that statutes contain definitions prefaced by "For the purpose of this statute. . . ."

-- Goals. The Board states its goals for students and for the process of schooling.

-- Accreditation. The Board here describes in detail its process for school standardization. That is, its process for responding to the legislative mandate to evaluate schools for compliance with the policy which is the major content of the Standards.

-- Instructional Planning. This section contains a rule which requires local districts to link evaluation and policy. This rule will be discussed in more detail below.

-- Instructional Programs. Here are the rules which express the Board's policies concerning the contents of a quality education program. These rules constitute the Board's response to the legislative mandate to establish standards of quality.

-- Administration. Rules for the operation of a school district. This section, coupled with the sections on Student Servoces, Staff and Class Load, Media and Materials, Facilities, Safety and Auxiliary Services describes all policy for the operation of a local district which is not contained in the Accreditation, Instructional Planning, and Instructional Program sections.

The purpose of examining the contents of the Minimum Standards was to see what the Board requires of local districts as a mix of policy and evaluation. It has already been stated that the accreditation section of the standards describes how the State Board will evaluate the local districts compliance with the Board's policies, so the accreditation section is a view of the Board's policy on how the Board will assure compliance, that is, how the Board will use evaluation. Our purpose is little aided by a detailed discussion of how the evaluation is conducted. The rules which describe this process appear in the appendix as 581-22-202, 204 and 206.

We are interested, here, in how the Minimum Standards have required local districts to use both policy and evaluation in the conduct of a quality school program. We are interested in the policy/evaluation interaction at the local level. Taking selected Standards, we will paraphrase or extract excerpts for the sake of brevity, and use these Standards to illustrate the interaction. The complete text of each Standard paraphrased appears in the appendix.

-- Standard 581-22-208, Instructional Planning. Each district is required to establish district, program and course goals. These goals embody the stated policies about the district's educational program and the outcomes of that program. Once stated, the goals form the basis of the local district's evaluation of its program and policies about education. This evaluation is required by the Standard. Following the evaluation, the district is required to identify its needs by comparing assessment results to its goals. The final requirement of the Standard is that, based on this evaluation process, the district is required to establish policies for making program improvements. Thus, we see that the State Board has directed districts to rely on the policy/evaluation interaction when planning its instructional program.

-- Standard 581-22-218, Educational Programs. The Board requires each district to:

(1) Identify individuals' learning strengths and weaknesses;
(2) Provide learning opportunities for students responsive to their needs;
(3) Determine progress students make in their educational program;
(4) Maintain student progress records and report the information to parents and students (OAR 581-22-218).

This required evaluation process is intended to achieve the Board's policy that districts "provide all students opportunity to achieve district-adopted learner outcomes, requirements for graduation and personal goals through participation in educational programs relevant to their needs, interests and abilities. (OAR 581-22-218)."

— Graduation Requirements. It is the State Board's policy that local boards "shall award a diploma upon fulfillment of all state and local district credit, competency and attendance requirements (OAR 581-22-228(1))."

Further, it is the State Board's policy that "student transcripts shall record demonstration of minimum competencies necessary to:

(1) Read, write, speak, listen;

(2) Analyze;

(3) Compute;

(4) Use basic scientific and technological processes;

(5) Develop and maintain a healthy mind and body;

(6) Be an informed citizen in the community, state, and nation;

(7) Be an informed citizen in interaction with environment;

(8) Be an informed citizen on streets and highways;

(9) Be an informed consumer of goods and services;

(10) Function within an occupation or continue education leading to a career (OAR 581-22-231(1))."

It is the State Board's policy that each local district expresses its policy about what a competent graduate is: "The local board shall . . . adopt and make available to the community minimum competencies it is willing to accept as evidence students are equipped to function in the society in which they live (OAR 581-22-231(2))."

The State Board policy, then, requires the local district to evaluate each student to see if the student has achieved sufficient competence to be awarded a diploma. "Each local district enrolling students in grades 9 through 12 shall implement the competency component of its graduation requirements as follows:

(1) Establish minimum competencies and performance indicators beginning with the graduating class of 1978;

(2) Certify attainment of competencies necessary to read, write, speak, listen, analyze and compute beginning with the graduating class of 1978;

(3) Certify attainment of all competencies beginning not later than with the graduating class of 1981 (OAR 581-22-236)."

We have seen that the State Board's policy is that Oregon graduates be competent citizens. To achieve this, the Board requires each district to state its policies concerning competent graduates and the Board requires an evaluation of students as a vehicle for effecting its statewide policy of competent graduates. Oregon differs from some states in that, in Oregon, evaluation of the graduates is conducted locally and is uniform only to the extent that the general competency areas have been defined. In intent, Oregon is no different from any state which establishes uniform exit requirements and conducts a uniform statewide evaluation of all students.

Additional examples are possible, but the intent was to show that it is the policy of the Oregon State Board of Education that (1) policies be established and communicated, (2) evaluation be conducted to be sure these policies are implemented, and (3) insures that policy and evaluation interact by requiring decisions (about programs and students) based on that interaction. Here, then, we see that the content of the minimum standards demonstrates that, in Oregon, evaluation is required by policy and policy and evaluation interact in the decision-making process. We now turn to the second stream of our discussion of the Minimum Standards to see how the development of the policy is affected by evaluation.

The Role of Evaluation in the Development of Policy

We have seen that the Board's policy in the form of the Minimum Standards requires districts to use evaluation information when setting policy and when making decisions. It has been asserted that the policy about evaluation-based policy making in the state agency is the same as the policy illustrated by the excerpts from the Minimum Standards. It can be demonstrated that the Board and the Superintendent rely on evaluation information when establishing policy. Again, many illustrations are possible, but it seems reasonable to show that the policy statement which requires evaluation-based policy making (the Minimum Standards) is, itself, policy based on evaluation.

When the revision of the current (1976) Minimum Standards began in 1978, the Superintendent and the Board wanted information about the objectives of the standards and about implementation problems. In addition, they wanted information about the results of the implementation of the 1976 Standards. A study was commissioned and conducted by the faculty of the University of Oregon. The study was designed by the present author. In order to show the reader a real example of evaluation in the real life of the Department, the following discussion is based on the major portion of the original paper which presented the design to the Superintendent and the Board. This provides an insight into the realities of conducting evaluation for use by lay policy makers.

Evaluating Oregon's Minimum Standards

A Context in Which to Evaluation

Evaluation is the collection of information for the purpose of making decisions. The evaluation of a program such as the Oregon Minimum Standards consists of the collection of information for the purpose of deciding whether the Standards are effective as they are or whether change is needed. The information to be collected relates to goals, implementation procedures and observable results. None of the information, however, is useful unless there is a well-defined model for program development and evaluation known to those involved in the evaluation so that the collected information can be "inserted" in the proper place in the decision making process. Simply put, if we don't know how we wish to use collected information (i.e., what decisions we wish to make), we will not know what information to collect nor will we know how to use whatever information we do collect. Before we can begin a discussion of the information we want and how we plan to get it, we must view the context within which this evaluation should be conducted.

There are many planning models ("planning" as used here subsumes evaluation and the making of policy). The model presented to the State Board of Education and adopted tentatively as the Department's generic model appears in Figure 3.* A simplified version appears in Figure 4. This model for planning and evaluating programs and policy, briefly described, requires:

1. The development of GOALS;
2. The identification of NEEDS (by comparing "what we want" (GOALS) to "what is");
3. The identification of LONG RANGE OBJECTIVES;
4. The identification of short term PROGRAM CHANGE OBJECTIVES which, when achieved, will move us toward the achievement of the GOALS;
5. The development of a PLAN to achieve some of the PROGRAM CHANGE OBJECTIVES;
6. The implementation and eventual evaluation of the effectiveness of the PLAN (an effort which seeks the answer to the question "are the goals of the PLAN being achieved?"); and finally;
7. We ask the questions "are the PROGRAM CHANGE OBJECTIVES, LONG RANGE OBJECTIVES and original NEEDS being met by this plan?" Following this effort we make decisions about the efficacy of the PLAN and the policies.

If an evaluator who knew nothing of the history of the development of the Minimum Standards were given this generic

---

*The Oregon Planning Model was developed jointly by Gordon Ascher, Robert Clemmer and Donald Egge, all of the Oregon State Department of Education.

Figure 3: Oregon Generic Planning Model

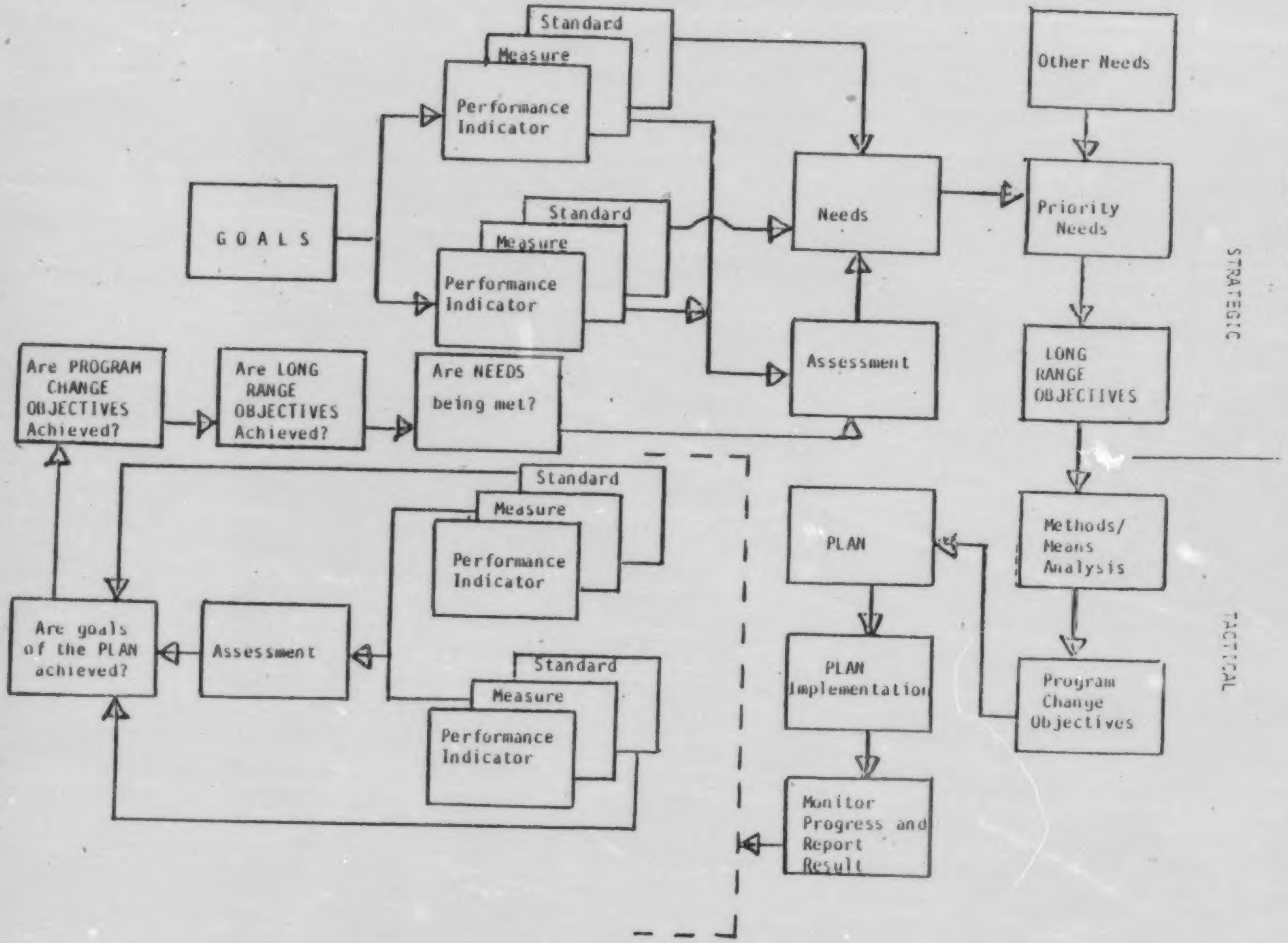
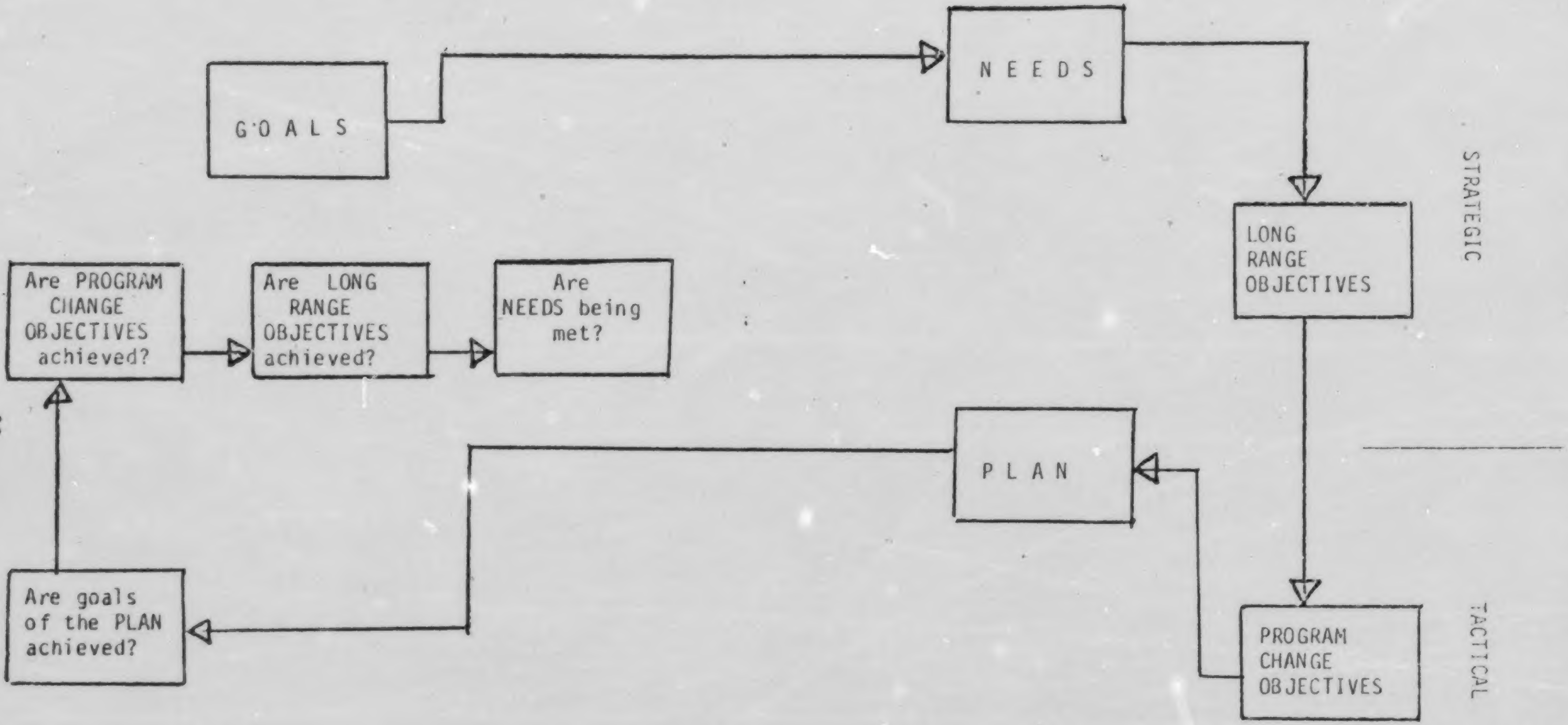


Figure 4


GOALS
NEEDS
LONG RANGE OBJECTIVES
STRATEGIC
PROGRAM CHANGE OBJECTIVES
TACTICAL
PLAN
Are goals of the PLAN achieved?
Are PROGRAM CHANGE OBJECTIVES achieved?
Are LONG RANGE OBJECTIVES achieved?
Are NEEDS being met?

model and a copy of the Minimum Standards, he/she would probably assume that prior to the adoption of the Standards, the Board:

1. Established GOALS which described the Board's philosophy and what it hoped to accomplish through a variety of means;
2. Used assessment procedures to collect information which, when compared to the GOALS, identified the NEEDS (i.e., the Board used the discrepancies between what it wanted and "what was" as statements of NEEDS);
3. Identified some LONG RANGE OBJECTIVES;
4. Identified some short term PROGRAM CHANGE OBJECTIVES which, if accomplished, would lead the Board closer to the achievement of its GOALS;
5. Developed a PLAN (the Minimum Standards) which it hoped would achieve some of the PROGRAM CHANGE OBJECTIVES, and
6. Implemented the PLAN with an evaluation design to answer the question "Are the goals of the PLAN being achieved?" All of this so that, within the context of the generic model, the Board could appraise how well the identified NEEDS were being met.

The evaluator's assumptions would be reasonable but erroneous. In fact, prior to the implementation of the Minimum Standards, the Board did not establish GOALS, identify NEEDS or identify LONG RANGE OBJECTIVES or PROGRAM CHANGE OBJECTIVES. The Board did develop a PLAN, albeit in the absence of a sound context, but neglected to develop an evaluation component for the PLAN. As a result, the Board cannot, it would seem, readily conduct a proper evaluation, after the fact. But, almost all of the parts of the generic model do exist and were attended to but these parts were not brought together as a cohesive whole. It is possible to do so because of the vantage point presented by the passage of time and because the work done to date has been of high quality.

We can conduct a proper evaluation because it is possible to reconstruct some of the missing links in the generic model. The evaluation answers four basic questions, "Are the PROGRAM CHANGE OBJECTIVES being achieved?"; "Are the LONG RANGE OBJECTIVES being achieved?"; and "Are the identified NEEDS being met?"

The Need for Reconstruction

Considering the question "Are the PROGRAM CHANGE OBJECTIVES being achieved?" To answer this question we need to know what the PROGRAM CHANGE OBJECTIVES are* that we are trying to achieve through the implementation of the Minimum Standards (the PLAN). Further, to evaluate the effectiveness of the PLAN we need to know the goals of the PLAN itself, and the performance indicators, measures and standards related to each of the goals of the PLAN. Of all of these, we have only the PLAN. We must identify the PROGRAM CHANGE OBJEVTIVES, the goals of the PLAN and the performance indicators, measures and standards related to each goal.

Once we have answered this question we can ask whether the LONG RANGE OBJECTIVES and NEEDS are being met. To do so, however, we must identify the LONG RANGE OBJECTIVES and NEEDS and to do that we must first identify the GOALS and the assessment information which enabled us to determine needs relative to the GOALS.

Upon accomplishing all of this, we will have the elements of a proper evaluation.

A Plan for Conducting the Evaluation

Part I. Evaluating the Minimum Standards Out of Context

This section presents a plan for answering the question, "Have the goals of the Minimum Standards PLAN been achieved?" As point out above, this question is important but it is only one of our important questions. Trying to answer the remaining three questions** requires placing the answer to the first question within the context of the Board's generic planning model. Evaluation of the Minimum Standards within that context is the subject matter of Part II.

---

*We could use the past tense but we want our results to be useful now so we will forget historical objectives and work with those currently in place.

To determine whether the goals of the Minimum Standards PLAN have been achieved we must first identify the goals and then identify performance indicators for each goal. Following that, we must identify appropriate measures and standards to determine whether the performance indicators were achieved and then make inferences about the achievement of the goals. We can define twelve tasks:

1. Identify the goals of the Minimum Standards PLAN.
2. For each goal, identify one or more performance indicators.
3. For each performance indicator, identify one or more appropriate measures and performance standards which will be used to collect and analyze information about the achievement of the performance indicator.
4. Collect and analyze data.
5. From information about performance indicators, infer about the achievement of each goal.
6. Identify the plan for implementation of the Minimum Standards.
7. Identify the implementation plan goals.
8. For each goal, identify one or more performance indicators.
9. For each performance indicator, identify one or more appropriate measures and performance standards which will be used to collect and analyze information about the achievement of that performance indicator.
10. Collect and analyze data.

**The remaining three questions are: (1) "Have the PROGRAM CHANGE OBJECTIVES been achieved?", (2) "Have the LONG RANGE OBJECTIVES been achieved?", and (3) "Are the NEEDS being met?"

11. From information about performance indicators, infer about the achievement of each implementation goal.
12. Produce a report on the achievement of program and implementation goals.

Part II. Evaluating the Minimum Standards Within the Context of the Generic Planning Model

Here we attempt to answer the questions "Have the LONG RANGE OBJECTIVES been achieved?" "Have the PROGRAM CHANGE OBJECTIVES been achieved?" and "Are the NEEDS being met?" To answer these questions we must first have the answer to the question posed in Part I "Have the goals of the minimum Standards PLAN been achieved?" Before we can answer these three questions we must also fill in the missing parts of the generic planning model. These missing parts are the State Board of Education GOALS, identified NEEDS LONG RANGE OBJECTIVES and PROGRAM CHANGE OBJECTIVES. We can define six tasks:

1. Identify the State Board of Education GOALS.
2. Identify the NEEDS.
3. Identify the appropriate LONG RANGE OBJECTIVES.
4. Identify the appropriate PROGRAM CHANGE OBJECTIVES.
5. Apply the data garnered from the evaluation of the Minimum Standards PLAN (Part I, above) to:
   a. determine whether the PROGRAM CHANGE OBJECTIVES were achieved,
   b. determine whether the LONG RANGE OBJECTIVES were achieved,
   c. determine whether the NEEDS were met,
   d. infer about the achievement of GOALS.
6. Produce report on the achievement of LONG RANGE OBJECTIVES, PROGRAM CHANGE OBJECTIVES and the meeting of NEEDS.

34

Results of the Study

The preliminary portion of the evaluation study has been conducted. The results of an extensive survey of several school-based populations are available. These results are currently being used by the Superintendent and the Board in the setting of new policy.

Summary

This paper has explored a conceptual framework for the relationship between policy and evaluation and has provided examples of the policy/evaluation interaction in the Oregon State Department of Education. The author has attempted to show that policy makers are very much dependent on evaluation information and, in fact, are required to be so dependent by the policies of the State Superintendent of Public Instruction and the State Board of Education. In addition, it has been shown that evaluators and the nature of evaluation are guided by the Department's written policy.

THE RELATIONSHIP BETWEEN EVALUATION AND POLICY IN THE VIRGINIA STATE DEPARTMENT OF EDUCATION

By

Gerald W. Bracey, Director
Research, Evaluation and Testing
Virginia Department of Education

Some General Comments and Definitional Issues.

Judging from the literature I have read, evaluation and policy appear to have been only tangentially or occasionally related at either state or federal or even local levels. Where they have been related, the relationship has generally been unidirectional with policy having a significant bearing on evaluation, but not vice versa. Explanations for this phenomenon vary. Fox (19), reviewing 10 years of evaluation of Teacher Corps concluded that the problem lay primarily with the methods chosen. According to Fox the "Standard Experimental Design" as defined by Astin and Panos (1971) was often used and was invariably inappropriate. Fox recommended to evaluators the "model" by Parlett and Hamilton (1976) described as "Evaluation as Illumination". Certainly evaluators have been misled by certain methods. However, in this case, I feel that a careful reading of Fox' paper leads one to conclude that the policy changes affecting by Teacher Corps would have taken place no matter what evaluation model had been chosen. Policy changes were determined by other factors and the outcomes of evaluations were simply irrelevant to such decisions. Indeed, House (1979), has argued that the Follow Through Program in particular and federal programs in general have been evaluated in such a way that the evaluations are bound to be narrow, trivial and hence irrelevant. Many would argue, I included, that the current Title I evaluations will be another case of a disfunction and disjunction between evaluation findings and policy changes.

All of this prologue is simply to document that the relationship between evaluation and policy has not been a happy one, at least as viewed from the perspective of the evaluators. Alkin and Daillak (1979) recently lamented:

> There have been great hopes for evaluation, not only among evaluators themselves, but also among many other educators, elected officials and the public. Yet these hopes have dimmed. It was hoped that evaluation information would help planners, administrators and policy makers both by improving individual programs and by

---

Cautionary note: While readers may experience difficulty with this paper for a number of reasons solely the fault of the author, readers addicted to the words "implement" and "impact" will experience additional difficulties as these words occur nowhere in the text in either noun, verb, or adjective form.

> aiding in choices among programs. The reality we are told by almost all observers, is that evaluation has had little influence on educational decision making, and evaluation information is largely ignored.

Although most evaluators (and even policy makers) reading that statement would probably give general assent, the statement by itself has several problems. It does not delineate types of evaluation or define policy. A necessary first step in understanding the role between "evaluation" and "policy" in the Department of Education or anywhere else is to define both evaluation and policy.

At one level any decision that determines policy is based in part on information that could be called evaluative. However, such information could be (and often is) derived largely from personal experience. Pipho (1978) reported that 80% of all legislation introduced with respect to Minimum Competency Testing Programs resulted from the immediate experience of a legislator with either his own family or that of neighbors or relatives. This datum alone should give some indication of the relative force of large evaluation projects on policy decisions. While these legislators are clearly evaluating information by using personal experiences as a source of evaluative information, such experiences do not constitute an adequate definition for evaluation for the purposes of this paper.*

For the purposes of this paper, "evaluation" or "evaluative information" will refer to any of the ten categories of studies classified by Webster and Stufflebeam (1978) as auasi evaluation or true evaluation.+ While it would be fascinating to include a discussion of what those authors refer to as pseudo-evaluation--propagandistic studies designed to support predetermined policies--such discussion would be difficult as, by definition, the goals are predetermined and part of the relevant information is either not collected or deliberately concealed by the perpetrators of the study. Note that one of the ten categories itself is labeled "policy studies" which foreshadows what will become clear--that the relationship between the two concepts is sometimes intricate and simultaneously elusive.

If one can obtain a frame of reference for "evaluation" by noting a single synoptic paper, such is not the case with "Policy". Attempts by the author to obtain concrete, short definitions of policy have failed.

---

*While personal experience does not constitute "evaluation," an evaluator with an interest in influencing policy would be well advised to see that his information became a part of the "personal experience" of the policy maker. Constraints on evaluators doing this will be discussed later.

+These categories are objectives based studies, accountability studies, experimental research studies, testing programs, management information systems, accreditation/certification studies, policy studies, decision oriented studies, consumer oriented studies, client oriented studies, and connoisseur based studies. Although Stufflebeam and Webster present them as "types" the author does not presume them to be mutually exclusive.

The general thrust of the answers has been "I can't define policy but I will know one when I see one". Dictionaries are no help and articles about policy provide little guidance because the word is used in so many contexts. It is endowed, as philosophers of science put it, with a great deal of "surplus meaning". Generally, people responding to the question, "What is Policy"? discussed an overall philosophy, ideology, or plan of action, sometimes with clear budgetary implications, sometimes without.

After a number of these discussions I imagined a continuum running from broad-scoped statements of "policy" at one end to specific statements of administrative procedures at the other. I was, however, unable to determine any point along that continuum where policy clearly became procedure or vice versa. Indeed, it will be seen during this paper that one source of difficulty in relating policy and evaluation is a confusion among all the actors involved as to where policy ends and administration begins. A result of this confusion is that administrative procedures, best left as such, get elevated to the level of policy making, with a resultant territorial intrusion by one group on the other. Similarly, policies get enacted with no clear procedural implications. Almost any course of action can be restated as policy. "It is the policy of the United States to contain communism wherever it should appear". Few would argue that this is not a policy of the U.S. government, but its procedural implications are by no means clear; the possible translative strategies range from providing effective demonstrations of the superiority of a capitalist system to providing assistance to any government promising to fight against communist activity in that country to physical obliteration of all nations professing themselves to have communist governments.

In other cases to be discussed, it appears that policy statements often make an appearance to justify administrative procedures already in place. Unless, of course, those procedures themselves are under attack.

Finally, in some cases the policy statement is redundant with or a gratuitous addition to a procedure. The 1978-80 Standards of Quality for Public Schools in Virginia, to be discussed in detail below, illustrate such a case of redundancy. One part of one standard states, "It is the policy of the Commonwealth that the awarding of a high school diploma shall be based upon achievement". There follows the standard that puts into place Virginia's graduation competency testing program. The policy statement is unnecessary, is unattached in any causal way to the testing program. In recognition of this fact, some legislators unsuccessfully attempted to have the policy sentence removed before the standard was enacted into law.

It is comforting, if not illuminating to note that others have struggled with definitions of policy issues and it is worthwhile to examine the struggles of others with policy.

Berlak (1970) pointed out that evaluators needed to know whether they were operating in an area of programmatic impact*or policy impact* and act accordingly. For Berlak, a policy issue has four criterial attributes:

---

*These are phrases from Berlak and do not violate the promise of page 1.

1. It directly or indirectly alters the power relationship between the citizens and the state.
2. It affects immediately or in the long run the status a person has and the power he can exercise within the social system.
3. It increases or decreases political or sicial tensions as a primary outcome.
4. It alters the self concept of the individual.

While not a model of clarity, nor in several instances easily amenable to empirical tests, these considerations are an improvement over the classical political definition of policy as the "application of reason and evidence to choose among program alternatives". Perhaps the best discussion of policy is by Mann (1975) who chooses like Berlak not to define policy but to describe policy issues in education again in terms of criterial attributes. For Mann, policy issues have five characteristics.

1. Policy problems are public. This might at first seem unnecessary to state as we are dealing with public education and hence all things are in public domain, but in fact they are not. Certain "policy" discussions revolve aroung what is the proper goal of public education. Which of the desired outcomes of childhood are the domain of the school and which are those of family and other institutions? Certainly much heat, if not light has been generated over where the responsibility of schools ends.
2. Policy problems have important consequences. They increase tensions among political groups and their resolution directly affects the lives of a large number of people or a small number of people in large ways.
3. Policy problems are complex. They have political, economic, moral, dimensions which of course, do not operate independently but interactively.
4. Policy problems involve large amounts of uncertainty. This almost follows from #3 above. If the state decides to allocate more funds to school districts with low scoring children, in order to hire more teachers what will be the outcome? Clearly this cannot be known in advance although various "scenarios" can be depicted with greater or lesser sophistication.
5. Policy problems are viewed differently by those with different interests or ideologies. Again this appears trivial or at least axiomatic but it is important to state. If all people agree on what is to be done, then it is no longer a policy problem. And, in terms of this paper, the fact that people disagree has direct bearing on how, when or if evaluative information will be used.

Attribute five, even if axiomatic, is important to the thrust and tone of this paper. I take it as axiomatic that readers will not be interested in the effect of evaluation on policy unless that evaluation contributes to the resolution of a policy issue.

The Structural Context.

There are, fundamentally, three policy making agencies in the area of education:

1. The Department of Education, a part of the executive branch of the government.
2. The State Board of Education, appointed at the pleasure of the Governor and operating through the Department, but often quite independent of it.
3. The legislature.

Policy matters may originate with any of the three. Generally, any policy matter originated in the Department is brought before the Board for approval, and the Board, seeking to take a stance on a policy matter, may instruct the Department to provide the information necessary to making decisions.

I will show in some detail in the next section that most major policy issues either enter the arena via the constitutionally required "Standards of Quality for Public Schools", also to be discussed below, or such policy issues come to rest there. In general, the Standards of Quality (SOQ) are drafted each biennium by the Department staff, approved with revisions by the State Board and approved with further revisions by the legislature. The legislature enacts the SOQ into law. It is thus obvious that the legislature is the final arbiter of educational policies contained in the SOQ no matter where the policy originated (unless the entire act is vetoed by the Governor). Again, as I will discuss in both the Historical Context and the Case History, the legislature in recent years has been very active in resolving new policy issues and reacting to policy matters brought to them by the Board and the Department.

The Governor may, of course, independently formulate policy through his office or through those of the Secretary of Education or the Superintendent of Public Education. In recent years, where Governors have acted independently, the target of their actions has been largely the domain of post-secondary education, not elementary or secondary public schooling. The Superintendent of Public Instruction and the members of the Board are all appointed by the Governor, subject to approval by the legislature. According to the constitution the Superintendent's term is coincident with that of the appointing Governor. In fact, no Superintendent has gone out of office with the appointing Governor. It might be noted the Virginia governors may hold two four-year terms, but not successively. The nine members of the State Board are appointed for a four year term and may succeed themselves once. Terms are staggered such that no more than one vacancy occurs at a time. In the two and a half years that I have been with the Department, five of the nine members have been succeeded, although one of the five resigned due to the press of other duties as the mayor of a large city.

The entire current Board have been appointed by Republican Governors although approved by Democratically controlled legislatures. It is sometimes alleged that the differing views taken by the Board and the legislature on matters in recent years are a reflection of this differing partisanship affiliation. That the legislature fights some of its battles with Governors through the vehicle of overriding the Board. Not only is the truth of this statement not verifiable, the allegation is not readily apparent in public exchanges. I was told by several people in all seriousness that the differences produced by partisan affiliation would be obliterated by the communality engendered by Virginia's longstanding commitment to conservative tradition.

Finally, the media, especially the print media, have shown no reluctance to propose or react to policy issues relating to public education. This might be expected by the description of policy issues in the preceding section. As pertains to evaluation, the press has been particularly active in commenting on the results of state mandated testing programs and trends in test results both in the state and nation. As will be seen, testing has come to constitute the bulk of evaluative information about public schooling at the state level.

Figure 1 shows the structure of the Department of Education as of December 1979. Only those areas outside of the Director of Research, Evaluation and Testing which are also concerned in significant ways with evaluation are labelled. Both the components of the Division of Research, Evaluation and Testing (DRET) and the way in which information flows from it to other parts of the agency merit comment.

It is an administrative regulation of the Superintendent that any information intended for the State Board must go through the Superintendent's office. While the previous Superintendent delegated much autonomy to the assistant superintendent for the flow of such information, the incumbent personally "signs off" on all such information. The importance of the quality of information flow from lower to upper levels in the hierarchy is manifest.

Similarly, it is a regulation that contacts between the Department, officially a part of the executive branch, and the legislature cannot be initiated by Department members. The Assistant Superintendent for Administrative Field Services serves as the official liaison between the Department and the legislature. Any information thought important to the legislature must be funneled through him. Legislators may contact members of the department informally for information or request them to testify before the various committees and subcommittees of either house. A brief report of all such contacts initiated by legislators in this way must be filed with the Assistant Superintendent for Administrative Field Services.

Virginia's DRET contains several functions which might not be expected to be found in a so-named group and does not contain several other functions which one might expect it to have. Notably it has no responsibility

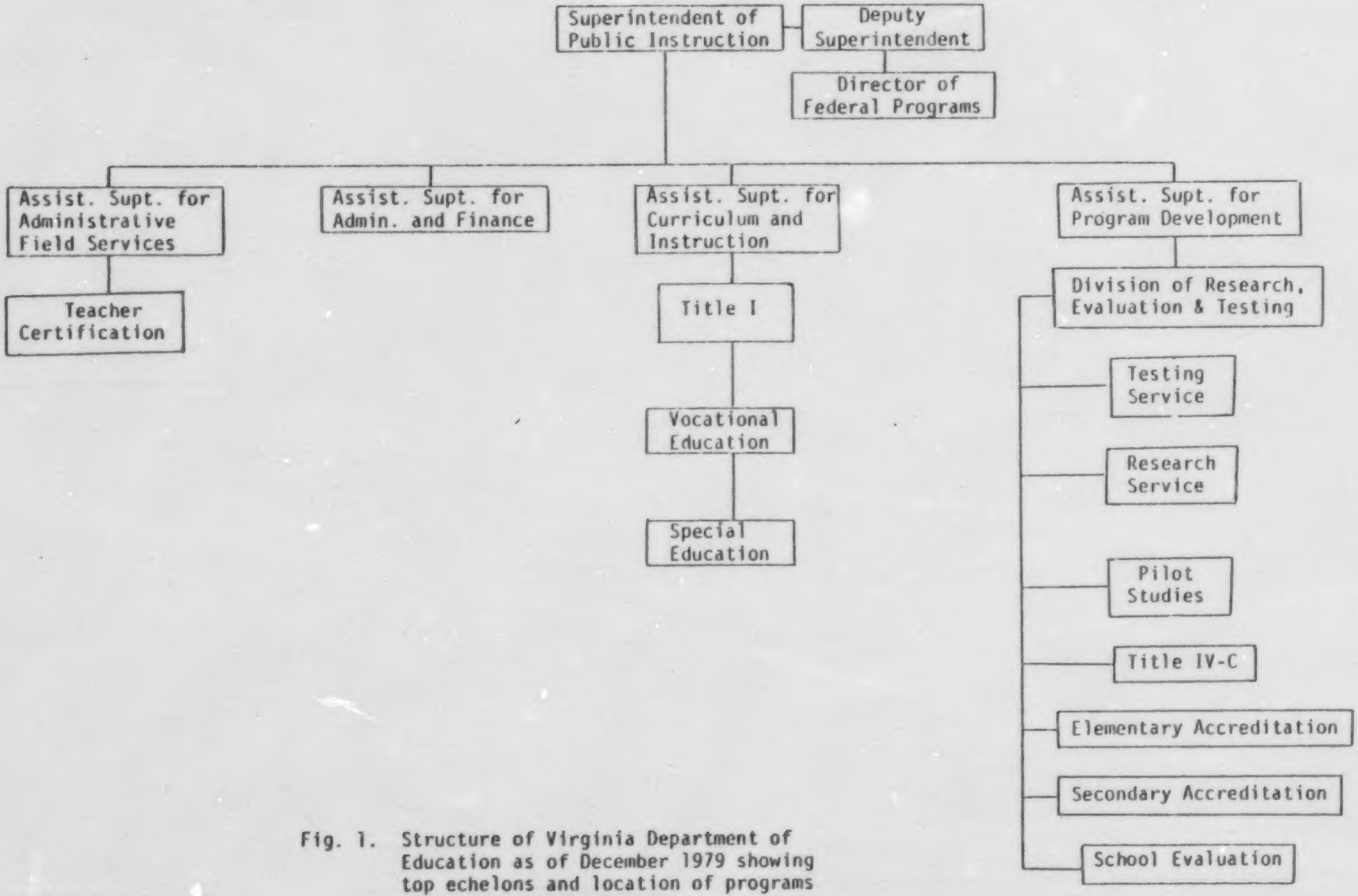

Superintendent of Public Instruction
Deputy Superintendent
Director of Federal Programs
Assist. Supt. for Administrative Field Services
Teacher Certification
Assist. Supt. for Admin. and Finance
Assist. Supt. for Curriculum and Instruction
Title I
Vocational Education
Special Education
Assist. Supt. for Program Development
Division of Research, Evaluation & Testing
Testing Service
Research Service
Pilot Studies
Title IV-C
Elementary Accreditation
Secondary Accreditation
School Evaluation

Fig. 1. Structure of Virginia Department of Education as of December 1979 showing top echelons and location of programs concerned with evaluation.

for the evaluations of Title I, special education, or vocational education. While this leads to administrative akwardness and some redundancy in efforts (children tested for Title I are retested under the state program, but the magnitude of any practice or regression effects is unknown) it has little real impact on policy at the state level: most of the above mentioned programs are constrained in terms of evaluation by federal, not state policies. Virginia has a long history of discordant relations with Washington and one often gets the impression that where federal funds are involved, the Department prefers the offices handling these tainted monies to be as autonomous as possible.*

This general feeling does not hold true entirely, as witness the placement of the Title IV-C office with the DRET area. While arguments could be made for its being elsewhere (and until 1977 it was with the Special Assistant for Federal Programs), coupled with the Pilot Studies program, which is a state funded innovative R & D effort, its placement with the Program Development sphere and R & D unit also makes sense.

What is most noteworthy and important is that while there is a research staff designated as such and a testing staff designated as such there is no evaluation staff designated as such. While accreditation falls into a valid category of evaluative information, the methods used in Virginia do not at the moment conform to the typical methods cited by Webster and Stufflebeam for these operations, notably self-study and team visits.

Accreditation is based largely on self-report on a questionnaire augmented by site visits. There is a secondary school evaluation designation, with one person in this role who organizes those teams which make site visits following a period of self-study by the local agencies. There are recommendations made from these committees but there is no follow-up to determine if the recommendations have been acted on and in any case no sanctions are imposed if problems are noted but recommendations are not followed. Virginia is one of few states which separates the accreditation and evaluation process for schools; the evaluation process is voluntary--it must be requested by the local superintendent.

The structure of the DRET leaves little or no opportunity for what is usually thought of as program evaluation. The research staff have skills appropriate to such evaluation, but for the most part their energies are devoted to either the periodic surveys that the department conducts or to assisting the testing service with the myriad of programs which have been mandated by the legislature in the past three years.

The lack of evaluation staff constrains the DRET in two ways: It must either rely on the assistance of people from outside the division not well trained in evaluative techniques or more likely, it must turn to an outside agency such as a university, contracted for a particular task. While such contracts have advantages as well as disadvantages, the chief disadvantages are that university faculties are often not aware of the information needs of the Department, try unsuccessfully to fit the problem into the paradigm of academic research, and cannot be "on site" often enough to render assistance at its most timely occasion.

---

*The Superintendent from 1975-78 was previously a local superintendent of the only division in the State that did not have a Title I program.

The Historical Context.

As indicated in the previous section, the use of evaluation is affected significantly by the structural context of policy making. It would be a mistake, however, to view this context as static. Indeed the purpose of this section is to treat that structural context as a dynamic, fluid one and show how it has changed over the past decade. The primary means of focusing on the historical changes will be by following the evolution of the Standards of Quality for Public Schools in Virginia and the evolution of the position of Director of Research, Evaluation and Testing as reflected, in part, by changes in the job description for that position.

While any starting date for a "history" would be arbitrary, the year 1971 seems appropriate to demarcate a formal change in thinking about education. By 1971, the policy of massive resistance to school integration had in itself been largely abandoned. No issue had dominated public education in Virginia quite the way that the ramifications from Brown vs. The Board of Education had ("This will keep us in power for at least twenty-five years", said one legislator in the 1950s, clearly seeing how long the issue would be in the forefront), but in 1971 the debate revolved not around whether, but how (i.e. through voluntary efforts or court-ordered busing). More importantly, in 1971, the Commonwealth of Virginia approved a new constitution for the state. A part of that constitution reads as follows:

> Standards of quality for the several school divisions shall be determined and prescribed from time to time by the Board of Education, subject to revision only by the General Assembly.
>
> The General Assembly shall determine the manner in which funds are to be provided for the cost of maintaining an educational program meeting the prescribed standards of quality and shall provide for the apportionment of the cost of such program between the Commonwealth and the local units of government comprising such school divisions. Each unit of local government shall provide its portion of such costs by local taxes or from other available funds. ART/VIII #2.

The Standards of Quality (SOQ) have come in the last eight years to be the final resting place of all policy issues where some state agency is the initiator of policy. Almost all major policy matters have either begun with changes in the SOQ, or if begun elsewhere (as, say, a change in the standards for accrediting secondary schools), have eventually made their way into this document. Similarly, most of the debates over policy issues at the state level can be found reflected in the changes of the SOQ over time and in the manner in which they were changed by the respective bodies who write and rewrite them.

The original SOQ for 1972-74* were performance-oriented with a number of standards being expressed in quantified terms relating to outcomes for both schools and individuals. Most of these outcomes were countable and stated in terms of expected changes in the future. For example:

> At least 45,000 five-year-old children in the State will be enrolled in kindergarten (26,500 in 1969-70).

Only one standard actually dealt with learner outcomes, and that standard, meaningless on its face was to cause considerable controversy and lead to changes in the SOQ. This standard and the controversy it produced will be discussed as part of the illustrative Case History.

In 1973, the position of Director of Program Evaluation (later to become Director of Research, Evaluation and Testing), was created and carried, in part the following job description:

1. To provide leadership in evaluating State objectives (Purposes of education adopted by the State Board of Education), programs (including standards of quality), and student progress.
2. To develop a training program for the State Department of Education staff and for school division personnel in translating state objectives into learner-oriented objectives, many of which should be measurable.
3. To develop, with the assistance of consultants and a representative committee, the criteria needed by school divisions to evaluate their own programs, organization and procedures, reporting and progress--especially the progress in student learning.
4. To assist schools and school divisions in making realistic evaluations and reports to the public.
5. To encourage and assist institutions of higher learning to evaluate their programs for pre-service and in-service preparation of teachers.

This broad, wide-ranging set of duties was never realized. The Program Evaluation staff consisted at first of only the Director and one other professional and later a second professional staff member was added. One reason for this lack of support was that the Standards of Quality themselves were being revised substantially for the 1974-76 version in a way so as to downplay the need for evaluation. In this version, the performance standards which had led off the list of original standards were eliminated. The general goals of public education were stated and five modified performance standards were listed as "objectives" with the note that "school divisions may wish to establish additional specific objectives to receive priority during during the biennium".

---

*Virginia operates most of its programs on the basis of biennial plans.

The bulk of the standards, listed as such, were ten prescriptive statements categorizable as "input" standards. They specified that schools were to provide kindergarten, special education, gifted and talented education, a certain number of professional staff for each certain number of students, etc.

This shift away from the outcomes of schooling back to the provision of goods and services--inputs--did not sit well with the legislature. According to persons I interviewed who were around at this time, the legislature felt that the intention of the SOQ section of the 1971 constitution had never been properly enacted by the State Board of Education. In any case, the 1974 session of the legislature noting that there had been "discussion both within and without the General Assembly as to what is appropriate to be included in the Standards of Quality...," created a joint subcommittee to study the SOQ. Formed in September 1974, the committee delivered its final report in December, 1975, just before the General Assembly convened for its 1976 session.

The words of the report itself reveal as well as anything can, the scope and philosophy of the subcommittee's study:

> Rather than confining its work only to the language of the standards, the joint subcommittee has sought to review comprehensively all aspects of publicly financed education in Virginia.

The outcome of this review, again in the words of the committee was:

> To a great extent, the Joint Subcommittee's revision /of the Standards of Qaulity/ has been based on the concept that the quality of education is measured ultimately by what students have learned (output) rather than the quantity or quality of resources devoted to education (input). Whereas some standards must be oriented towards input, the greater emphasis should be, in the opinion of the Joint Subcommittee, on output.

Thus very clearly and dramatically did the General Assembly, on adopting the SOQ as revised by the Joint Subcommittee change the thrust of educational policy away from the traditional goal of providing goods and services to measuring outcomes. In fact, many people outside of the General Assembly felt that the legislature had not revised the SOQ so much as they had actually rewritten them, thereby overstepping their constitutional powers. According to these same observers, the Board decided that any constitutional challenge could result, at best, in winning a battle but certainly losing the war; accordingly, no such challenge was made.

Two other quotes from the final report are worth noting because they set the stage for the introduction of two Standards which were indeed written, not revised, by the General Assembly. Continuing with its focus on outputs the Joint Subcommittee concluded its introduction with a set of premises, the first two of which are as follows:

1. The basic purpose of the Standards of Quality is to establish minimum elementary and secondary educational goals that are to be met for each child (to the extent practicable) throughout the Commonwealth.
2. Standards established by the General Assembly should be oriented primarily towards products (objectives, outputs and goals) rather than processes (inputs and means), thereby creating a structure and environment for quality education.

To the best of my knowledge this is the first use of the word minimum in connection with the outcomes of public instruction in Virginia and provides the basis for a wide ranging set of changes in the orientation of public education. With these premises in mind the General Assembly wrote the following standards:

> Standard 1A. The General Assembly concludes that one of the fundamental goals of public education must be to enable each student to achieve, to the best of his or her ability, certain basic skills. Each school division shall, therefore, give the highest priority in its instructional program to developing the reading, communications and mathematics skills of all students, with concentrated effort in...grades one through...six. Remedial work shall begin for low achieving students upon identification of their needs.
>
> Standard 1B. By September 1978, the Board of Education, in cooperation with the local school divisions, shall establish specific Statewide minimum educational objectives in reading, communications and mathematics skills that should be achieved during the primary grades and diring the intermediate grades.

And how was the General Assembly or the public to know if concentration was being focused on these basic learning skills and if individuals were to be receiving appropriate remedial work? By means of tests:

> Standard 7A. By September 1978, each school division shall primarily utilize testing programs that will provide the individual teacher with information to help in assessing the educational needs of individual students.
>
> Standard 7B. Beginning in September 1978, each school division shall annually administer uniform Statewide tests developed by the Department of Education to measure the extent to which each student in that division has progressed during the last year in achieving the specific educational objectives that have been established under Standard 1B.

Standard 1 codifies a back-to-basics movement for Virginia. Standard 7 provides perhaps the most ambitious, comprehensive program of diagnostic testing in history. If, that is, the standards are to be taken at face value. And while certainly these standards, conceived entirely within the General Assembly are not to be taken lightly--the objectives and commensurate testing program are at present in place--there is good reason to believe that the legislature was not fully aware of the implications of what it was doing. The committee had been advised primarily by one legislative aide, untrained in psychology or education. While that aide read a great deal of background research, and while the committee as a whole learned a great deal about testing, the final report of the Joint Subcommittee is a melange of the Zeitgeist, theory, errors and naiveté. As paper after paper was delivered by the legislative aide to the committee, "We began," in the words of one committee member, "to conclude that he knew what he was talking about when in fact what was being created was legislation by inundation--we were simply inundated with papers about objectives and criterion-referenced tests and so forth."

The aide had indeed "discovered" criterion-referenced testing and proposed it as the only reasonable alternative to current testing practices. The following quote from the final report of the Joint Subcommittee is revealing of the errors and naiveté alluded to above:

> Far too much emphasis in testing has been placed on how a group of students (a classroom, school, division, or state) compares relative to a "norm" group. Relative rankings bear no direct relationship to an absolute level of academic competency.
>
> Particularly with basic skills, knowledge is more absolute than relative. Thus, use of relative rankings or percentile scores masks any change in the absolute acquisition of skills or knowledge. The Educational Testing Service, which administers the College Entrance Examiniation has noted a steady decline over the last ten years in the absolute academic achievement of students taking its examination. "Norm referenced" tests do not show the decline that has actually taken place.

Such delightful confusion could be accepted if it did not accompany a document proposing a program that no one had yet accomplished--getting all or most teachers to use tests, and tests, as meaningful, not to say diagnostic instruments!

The Department of Education's involvement in this policy change was nearly nil. Indeed, the Department had been operating independently on its own initiative. The Director of Program Evaluation had convened in 1975 a State Testing Committee made up of LEA, SEA and IHE personnel to propose a comprehensive testing program for the state. Much, though not all of their work was rendered moot by the actions of the legislature (the State Testing Committee did not make its final report until December 1976, some nine months after the action of the legislature). The Department apparently had no inkling of what the legislature was about to propose; the legislature, for its part, did not acknowledge the existence of the State Testing Committee.

The effect of the new SOQ for 1976-78 was to strengthen an equation that had been in the making for the preceding two years. That equation was simply evaluation = testing.

The new SOQ were passed in March of 1976, the Director of Program Evaluation resigned in August and the position remained vacant for nine months. When it was advertised again, the job description had changed considerably. Now called the Director of Program Evaluation and Testing, the new ad specified a person to provide:

> Leadership in developing a program for measuring individual student attainment of basic skills; developing a program for assessing student achievement; coordinating the work of staff members in developing, administering, and interpreting the statewide testing program; and developing and administering the testing program budget.

Evaluation = school outcomes = equal test scores. A rather different orientation than in 1973.

It is quite possible that all testing except that prescribed by the new SOQ would have gone by the Boards in 1976 had not Virginia's intermittent policy making body, the press, jumped into the fray. In both articles and editorials, newspapers, particularly those in Richmond, the State Capital, argued that the elimination of norm-referenced tests would lead to chaos as it had in other states. California was cited as a state which had changed tests so often that no one knew where the state was, what the anchor for scores was. The NRTs were kept.

A Case Study.

THE EVOLUTION OF VIRGINIA'S GRADUATION COMPETENCY TESTING PROGRAM

If the history of the Standards of Quality provides a general framework for the evolution of the equation testing = evaluation, the history of Virginia's movements in the area of "minimum competency testing" provides a concrete example of that equation in action as well as how policy decisions are made and how programmatic decisions get elevated to policy levels.

In a period of two and a half years, Virginia moved from having no competency requirements at all, through a stage of having an "Oregon plan" with localities having latitude,to having a uniform statewide graduation competency testing program.

In 1976, acting to head off a legislative mandate and in reaction to "the handwriting on the wall", the State Department through the State Board proposed that a set of competency requirements be added to the standards for accrediting secondary schools. While some evaluative information was considered--test score declines as reported by ETS and NAEP and fluctuations in the results of the State Assessment Program--there is no evidence that the decision to add competency requirements was significantly influenced by these results.

The requirements dealt with computation, communication, social studies and ability to successfully pursue post-secondary experiences either in the marketplace or in higher education. The first three were adopted almost verbatim from a publication of NASSP entitled This We Believe. The fourth area was derived from the goals of public education as stated in the Standards of Quality.

In considering to whom this new requirement should apply first, the Department and the Board were guided largely by an intuitive sense of fair play. The Department and the Board felt that the children to be affected should have ample warning before hitting the barrier. In terms of the course of schooling, it made sense that the requirement could justly be imposed on those children who were about to enter the ninth grade. They would know about the requirement ahead of any high school years and have the full four years to meet the requirement.

The plan adopted was referred to as the "Oregon plan" although it differed significantly from the Oregon program in one respect: While Oregon allowed local divisions to establish the competency areas the State of Virginia specified them. However, as in Oregon, the determination of how to assess competency and how much of any competency constituted enough was left up to the localities.

The localities, with little input into this requirement, gave the action mixed reviews and treated it with mixed reactions. Some divisions already had a similar requirement and essentially ignored the Board's action.* Other divisions took the new requirement seriously and began to plan in various ways to meet it. The most common form of planning was to develop or purchase a test.

Almost immediately, a large number of local divisions began to lobby the Board, the Department and various legislators to move the competency assessment program to a statewide level. Some lobbying efforts derived from legitimate concerns, some did not. Many divisions in Virginia are small and lack the staff or money to develop or even purchase assessment devices in four areas. Others, it appeared, simply did not want the bother of the development process and some, seeing a fertile lode of future litigation, preferred that the state and not they be hauled into court. The pressures for statewide assessment of the competency areas grew during the fall of 1977 while legislative committees were meeting.

When the legislature met in early 1978, an addition was proposed to the testing and measurement standard created by the legislature in 1976. As originally written, the 1978 standard would have required tests to be used both for graduation and for promotion from grade to grade. This standard was debated hotly over a period of two months, but no member of the Research, Evaluation and Testing staff was called on to provide any kind of testimony about the wisdom of the standard. And, unlike the original testing and measurement standard which had been preceded by a year-long, if somewhat flawed study, no such study was conducted prior to the introduction of this new standard.

---

*There is some anecdotal evidence that some people did not really take the requirement "seriously" until after the first results of the later statewide tests were released in early 1979.

The addition of a standard calling for a statewide graduation test was really no surprise as the local superintendents had gone on record as favoring it. Calling for the use of tests to determine promotion, however was something of a shock. On what evaluative information was this statute proposed? One cannot be certain because of the lack of contact between the Testing Staff and the legislature, but it is reasonable to assume that the legislature was responding to what had become known as the "Greensville Experiment". To understand the Greensville Experiment and its power requires a digression, but it is an illuminative one.

The original Standards of Quality had called for a division's achievement level to match its ability level, both levels being determined by NRTs*. While most divisions had close correspondences between the two tests, Greensville County did not--the achievement level was well below the "ability". Based on these results, the State Board singled out for public censure the County of Greensville as failing to meet the Board's Standards. Greensville's response was to retain about 67 percent of the fourth graders and high percentages in grades 1, 2, and 3. All those who did not achieve at a certain level on the NRT were retained. While this caused a short term furor within the county, the superintendent was able to build public support for the program which involved more than simple retention based on test scores. Greensville's superintendent presented the program as offering low achieving children "more time to learn" and was able to convince a substantial portion of the population that it had this benign intent, thereby removing much of the usual stigma associated with "flunking". The superintendent was able to convince the majority of his constituents that it would be unfair to promote the children who had not achieved a certain level on the tests; that it would in fact be more humane to allow them more time to learn these skills by repeating the grade than to go on and encounter even more difficult subject matters in higher grades. By 1977 favorable reports on Greensville had been written in various state newspapers, Time magazine, and shown in prime time on two of the three commercial television networks. Greensville's program is complex and difficult to evaluate partly because of inadequate baseline data and partly because the need for administrative decisions which outstrip researchers' ability to gather data. Greensville is a small division with essentially a three-person central office. One can scarcely fault them for not being research oriented, for operating with an administrative style of deciding what needs to be done on the basis of the collective wisdom of the office and the school board and doing it. However, the general public and some legislators had gotten a simplistic notion of the program that said essentially, if you fail kids, scores go up. The fact that those children with low scores were being retained and that this selectivity alone would make scores appear to increase was not a part of this simplistic conception. Part, but only part, of what Greensville County had done was to test third graders locally and retain those scoring below a certain level. Thus when the succeeding fourth graders took the state required test at that grade, the fourth grade scores appeared to rise dramatically. Other increases occurred in other grades.

---

*In fact, as phrased, the standard did not have any meaning. The interpretation given to the standard was that the average percentile rank for a division in achievement should be equal to or above its average percentile rank for ability.

By the School year of 1977, however, the children retained in the lower grades for one or more years were now 14 years old. They were also in many cases, in ungraded classes. However, for purposes of funding by the state, the children had to be declared as either eighth graders or special education students. By declaring them as eighth graders they would be eligible for intermural athletics and they were so declared. But being now eighth graders, they came under the law that required all eighth graders to take the state NRT. The result was that the scores that were apparently high, took an equally apparent nose dive. A legislative resolution commending the superintendent of Greensville for his efforts was tabled and the promotion-by-test section of the proposed new standard was deleted. Such is the relationship between test data and policy. The standard calling for a graduation test was retained, however, and in its final form read as follows:

> It is the policy of the Commonwealth that the awarding of a high school diploma shall be based upon achievement. In order to receive a high school diploma from an accredited secondary school after January 1, 1981, students shall earn the number of units of credit prescribed by the Board of Education and attain minimum competencies prescribed by the Board of Education. Attainment of such competencies shall bee demonstrated by means of a test prescribed by the Board of Education.

Certain characteristics of this policy statement and its concomitant action requirement stand out. The policy is, in one sense extremely prescriptive in that a test is required. In another sense, the policy is extremely liberal in that no areas are defined for the testing. Under the letter of the law, the State Board is free to prescribe a test in the basic postures of hatha yoga. In one sense the policy is extremely vague in that it does not define test. And a number of people were concerned that because the standard referred to "a test" the Board was not free to prescribe different tests in more than one area.

Needless to say, hatha yoga is not a requirement for graduation and test has been interpreted as four-choice multiple choice. The Board decided that they could require more than one test--reading and mathematics--and no legislator has complained that this violates either the letter or the spirit of the standard. That such concerns about wording could be raised and discussed with a semblance of seriousness is indicative of the sometimes fragile relationship among policy making bodies.

While the legislature altered the policy of how competency was to be demonstrated, it left the year when the statute became effective unchanged. The class of 1981 was (and is) under the gun. Again, no one from the Testing Staff of the Department was asked to testify concerning the effect of these policy and action changes. At this point, no litigation had been resolved in Florida and McClung had only recently (1978) suggested that the length of phase-in time for such a program could be a source of litigation. The Department felt strongly, however, that the children to be affected should be tested as early as possible. Early testing would allow for remediation and to prevent any feeling that the rules of the game for getting a diploma had undergone a sudden and capricious change in the eleventh hour (or the eleventh grade). Accordingly, the Department convened two committees of reading and mathematics experts for the purpose of reviewing tests in these areas and recommending to the Board which should be used.

It should be noted in passing that the calling of these committees reflected a long standing Department policy. Whenever changes in educational programs are likely to have significant impact on local school divisions, committees representing different regions or different levels of organizational status or both will likely be convened to "evaluate" the change and advise the Department and Board. In my experience, such committees function substantively, not as window dressing.

The two committees consisted of teachers, supervisory staff, university faculty and representatives of various special interest organizations. Their charge was to review all tests that had been developed in Virginia under the "Oregon plan" years, as well as those tests that were coming onto the market from commercial publishers. Two tests were recommended to the Board and accepted by it in June, 1978.

With the tests chosen, another question arose: How much competency on these tests is enough? Where is the cut-score? This issue might seem at first to be too minor to constitute policy. On the other hand, it certainly meets all of the criteria for a policy issues as described by Mann and cited in this introduction. It is a public problem, with important consequences, with political, economic and moral dimensions, involving a goodly amount of uncertainty and viewed very differently by different interests. In any case, the amount of heat generated by the debate elevated the decision to the policy level.

Most of us directly concerned with testing were largely in agreement with Gene Glass (1977) that cut score decisions could not be based solely on technical considerations. We were, likewise in agreement with Glass that:

> For most skills and performances, one can reasonably imagine a continuum from "absence of skill" to "conspiciuous excellence. But it does not follow from the ability to recognize absence of the skill that one can recognize the highest level of skill below which the person will not be able to succeed (in life, at the next level of schooling, or in his chosen trade)... If anyone would dare to specify the highest level of performance below which no person could succeed in life as a parent, counter examples of persons reading below the minimal level yet who are regarded as successful parents could be supplied in abundance.

On the other hand, a cut score was necessary and thus the search had to be confined to methods that would reduce arbitrariness in its bad sense of being capricious. Several methods were considered and that which seemed most congenial to us was the method proposed by Jaeger (1978). While openly judgmental, the procedure allows for judgments by various audiences and a series of iterations before reaching a final decision which allows judges to change their minds based on new information. However, one of the cycles of the model requires that data from item field tests be provided to judges that they may not how many children actually answered a given item correctly. This cycle seems particularly important in view of other similar models which operate without actual data and often lead to unrealistically high scores (e.g. procedures for validating the National Teacher's Examinations).

However, with the legislation enacted in March and testing scheduled for November, clearly no field test data was going to be forthcoming and the two tests chosen had limited field testing and none in Virginia. In June of 1978, the testing staff recommended that the setting of the cut score be delayed until actual test results were back and that the results from the first administration be used in the Jaeger model where data was called for.

No action one way or another was taken on this recommendation for some months. In late August or early September the press became cognizant that no cut score would be established until results were in, and in several cities allowed as how such a procedure would permit the scores to be "tampered with" to make the results both politically and economically acceptable to the "educationist establishment". In reaction to these allegations in the press, the State Board attempted at its September meeting to set a cut score in the absence of any procedures. "Let's set a score and let the chips fall where they may," said the President of the Board. "We can always change the score if we need to". I and two other members of the department argued vehemently against this approach, and after several heated exchanges, won a month's delay to conduct some kind of study only by promising to provide a recommended score at the next Board meeting. During the month, seven groups of about 15 people each were convened in various divisions of the state at the invitation of the local superintendent of the division. The instructions to the superintendents were to pick people representing professional educators of all levels, and parents and other interested community members. Each item was examined in terms of its importance and then a global rating was obtained as to what would be a fair passing score. The range of scores was 35 to 85. When analyzed according to a lay-educator dichotomy, lay persons wanted a score around 75, educators around 70. The Department recommended 70 which was accepted by the Board.

It is worth noting that feelings among the testing staff were so high in opposition to this haphazard approach that others in the Department administration "absolved" us from any responsibility in the conduct of the modified, less-than-rigorous cut score study. In retrospect, I feel that this cut score was predetermined, that it was not likely to have been anything other than 70 (the judgment of the groups actually favored a score of 75 and had the testing staff been in charge of the recommendation, that would no doubt have been the recommended score; if the testing staff had conducted the study it is not clear where the cut score would have fallen).

Again, in retrospect, it was good that the cut score was in fact established before the test was given. Even though in the actual procedure political considerations and public relations considerations weighed more heavily than conceptual soundness, a technically sound study conducted before the test administration might have led to more "disruption and dislocation" in Glass' phrase and a cut score set after the test administration certainly would have mixinized disruption and dislocation. This unfortunate outcome would have been determined by events surrounding the test administration that could not have been foreseen prior to that administration.

The competency testing program had been good copy for the media. Both print and electronic media representatives had been keeping close tabs on events surrounding the competency program. When the test was actually administered, schools were deluged with reporters interviewing students for their reactions. The tests, said the students without exception, were "a piece of cake" and "an insult to my intelligence", etc. Apparently no reporter questioned whether or not his sample might be biased--that children who experienced difficulty would either steer clear of the cameras or be loath to admit hardship in the presence of peers sneering at the ease of the test. In at least one instance we obtained reports that only the "right" students were being aimed in the direction of the cameras by the school staff.

In addition, approximately a month after the test administration and before any results were received, sample items of both tests were released at a news conference. These items did little to convince the press that the test had been difficult and one paper published the items with the first of a series of vitriolic editorials attacking the test as a farce and public education as a sham, bilking the public by permitting children to graduate knowing so little.

One can only imagine what would have happened in this charged atmosphere if the cut-score had been set after all this publicity. It is likely that any group charged with establishing a passing score would have felt obligated to set the score higher than what had been established prior to the administration. With the score set at 70, 9% of all white students and 33% of all black students failed the reading test. If the score had been, say, 80, these figures would have risen to 25% and 63%, respectively.

Before leaving this particular topic, it should be noted that while the data preferred by the testing staff could not be gathered, some information was gathered for use in establishing a cut-score. Some 100 people did render judgments about the appropriate cut-score. The extent to which this data influenced the cut-score or which it "improved the decision making process" is undeterminable. What is clear is that this is the type of information that will often have to be used in informing policy makers. Because of time pressures, political, economic and public relations considerations, it is likely that the need for a decision will usually outstrip the professionally desirable methods for collecting information. If evaluators are not simply going to take their marbles and go home (thereby rendering their contributions nil), they must learn to cope with and use less than "pure" data. The techniques for such coping and use and the criteria for evaluating the power of such data are by no means clear, although the exponential increase in evaluation methodologies--everything now seems to be a metaphor for evaluation--is testimony that evaluators are at least aware of the problems.

With a cut-score set, the tests given, the question now arose as to how to release the data. For many years test scores were not released, but in 1971, the State Board, reacting to both public pressures for such scores on a division-by-division basis and an opinion of the Attorney General, decided to make such release a matter of course. In spite of an attempt to defuse invidious comparisons among the divisions by emphasizing the desired match between "ability" and "achievement" (noted on page 16), this practice did not sit well with many legislators as noted in the statements from the Joint Subcommittee report. Nevertheless, the Board's decision was not challenged and the Department planned a release in this fashion for the Graduation Tests. In addition, because of results already reported in Florida and North Carolina showing marked racial differences, it was decided to report the data analyzed by these two categories. The results were accompanied by a brief summary-interpretation prepared by the testing staff.

The racial differences in passing rates made big headlines around the state. In addition, a number of divisions that do not traditionally do well on the norm-referenced tests did very well on the Graduation Test, much better than neighboring divisions with similar NRT scores. This led in some instances to phone calls to the Department questioning the validity of the results for the high scoring divisions and in a few cases to open charges that some divisions had somehow cheated--had taught to the test, withheld certain percentages of their kids who should have been tested, etc.

Such accusations place the Department in a difficult position. It lacks the resources to administer the tests and certainly cannot verify the challenges of impropriety. The best that the Department can do is to prepare a narrative to defuse these kinds of attributions. However, this course has its own dangers. If the narrative is too long and makes too many points, the media accuse the Department of managing the data and ignore the narrative.

If the narrative contains information which the Department feels is important but which is complex--which cannot be dealt within a few paragraphs or during a 60-second segment on the evening news--such a narrative is also likely to be ignored. For example, while the failure rate for blacks was four times that of whites statewide, this rate did not prevail in all divisions. Indeed, in a few divisions, blacks passed at a higher rate than whites, and an examination of those divisions where blacks performed well indicated that the results could not be a simple function of demographic variables, SES or anything simple. In all likelihood they reflect subtle program and extra-school variables. But sublety has no place in the face of deadlines and short on-camera reports, although the variations discussed above were reported in the narrative.

The point of this discussion is that evaluators in possession of "public information" often have a difficult time getting that information to the public in usable form. The Department is often forced to do a "data dump" and hope for the best. Presentation of what the Department considers important is often viewed with suspicion as being that which the Department considers to be in its best interest.

I would note in conclusion, however, that in fact there is never really a "data dump", but only a method and format of reporting certain information to which the press and others have become accustomed. It has been noted by numerous philosophers of science that "data" should really be called "capta"--that nothing is given but rather is taken, captured. The problem for evaluators is to obtain acceptance for the kinds of capta they consider relevant, not simply those which may be relevant to the press.

POSSIBLE FUTURES FOR EVALUATIVE INFORMATION

In the preceding sections of this paper, we have seen that the influence of evaluation on policy in Virginia has been a very limited one, having been constrained by the structural context, as well as historical events. The structural constraints were two-fold: an absence of evaluation staff and strictures on the flow of information which would make it difficult to get relevant information to those making policy decisions. The Zeitgeist has also constrained the influence of evaluation in two ways. People have not seen evaluation as particularly relevant or important to policy except in a negative way. Secondly, the spirit of the times says that in fact, the evaluation data are already in: Test scores are declining, costs are going up, teachers cannot teach and the whole structure needs revamping. This aspect of the Zeitgeist clearly includes the general public, as can be seen in the annual Gallup polls on public opinion of education, as well as the policy makers as can be seen from the number of legislated educational programs.

In Virginia, one senses, though it is difficult to demonstarate with "hard" evidence, a punitive, calvinistic attitude towards the schools. The schools have failed to control the natural depravation of mankind and must be punished for their shortcoming. It is never phrased that way, of course, but the behavior speaks louder than the words. I have often been asked by people outside of Virginia what remedial programs the state is providing for those who fail the Graduation Competency Test. My answer is none, there is no money for remedial programs.

When the same enquirers have asked "Is this not a little unfair", my reply is this: Many who could provide funding believe that if the schools were doing their job there wouldn't be anyone failing the test and so no additional money will be forthcoming to assist what should be done anyway. Not all policy makers feel this way but enough do to prevent any sums being appropriated for such programs or for legislation being introduced to provide money.*

Similarly, the Standards of Quality can be assessed in terms of their per-pupil cost. That is, how much money on the average, does it cost to provide the programs required by the Standards of Quality. According to data collected at the local level and state level, the standards have consistently required more money than the legislature has appropriated for them. This failure to appropriate actual costs for constitutionally required standards, largely written by the legislature reflects again a certain attitude of punitiveness towards public education.

As another instance of constraints on the use of evaluation at the state level in Virginia, let us consider a situation in no way unique to Virginia, but rather symptomatic of certain organizational structures. Evaluation, to be used properly, must take place in an atmosphere where there is some freedom to fail. If one is not doing pseudo-evaluations where the results are determined beforehand, one cannot guarantee the outcome of an evaluation. Such freedom does not exist in education and such and such freedom does not usually exist in bureaucratic structures competing with other such structures for money and power. Rich (1979) distinguishes two different ways of "avoiding risk" in organizations. For scientifically oriented academic researchers, reduction of risk consists of and in the ideal requires, new information which may contradict earlier information. The goal is truth and the new information reduces the risk of being wrong. While I think Rich's description of academia's open arms acceptance of new information is self-serving and overstated#, certainly there is more freedom to fail in academic settings.

---

*After the first draft of this paper was written a part of the 1980-82 Standards of Quality was changed to permit the hiring of extra personnel for eighth-graders scoring three or more years below grade level on the eighth-grade norm-referenced tests given annually. These children are known to be at some risk in terms of passing the Graduation Competency Test. The change thus has the effect of providing remedial assistance although it is not phrased that way. This kind of indirect use of evaluation information is becoming more common in Virginia.

#The case of acupuncture vs. the American Medical Establishment comes immediately to mind. Rich's statement might be better phrased, to the effect that new information is welcome in direct proportion to its potential for getting grants and publications for the receiver of the information. Resistance to change in the scientific community has been beautifully and amply documented in Thomas Kuhn's The Structure of Scientific Revolutions.

Rich is more on target when he speaks of the Manager's Perspective of Risk Avoidance. Here risk relates to competition for scarce resources. The manager in a bureaucracy is likely to ask how new information might be used to embarass him or to place another agency in a more favorable light. Evaluation in this context does not have a "truth" orientation, but a political one. Evaluation will be well received if it helps the organization and manager meet goals, minimize costs and maximize gain. This means that evaluation as customarily defined has an uphill fight--it takes place in a value-laden political context. Education is in an area of competition for scarce resources and in an era of negative opinion towards its achievements. In such an area in such an era the cost of looking bad is too high to permit a scientifically detached perspective on evaluation.

Finally, the role of evaluation in Virginia is constrained because it has been narrowly defined, namely by test scores. It is safe to say that until test scores cease to be an area of concern, such potentially fruitful sources of evaluation as affective variables or process variables will be ignored. Similarly, unitl test scores per se become a non-issue, little will be done in the way of process or formative evaluation.

It would be easy after reading the above paragraphs to conclude that evaluation is unlikely to ever be usefully applied in Virginia through the Department of Education. Easy, but wrong. While events do not augur well, there are areas where imaginative and energetic use of data and lobbying by those interested in evaluation can produce results.

The present press for accountability in education is likely to increase. The problem for evaluators is to eliminate the equation of accountability with test scores and write a set of equations including many variables. While there is no guarantee that such a conceptual broadening can occur or be used properly, there is at least one area of accountability where it might. The focus on competency in Virginia has moved from students to teachers and while there is, yes, a test required for new teachers, all concerned recognized that such a test will not guarantee anything except the screening out of near-illiterates. There is a recognition that knowledge and competence are different and that competence is related to behavior. This is not a recognition well articulated yet, but it is there and the necessity to move towards assessing teacher performance would seem to at least open the door to a broader conceptualization of evaluation.

Similarly, evaluations could take policy issues and relate them to empirical research in such a way as to broaden perspectives and hopefully open the way for better utilization of information. For example, class-size has been a policy issue in Virginia for some years. There has been a mandated reduction in class-size as apart of the SOQ for some years. The mandate occurred without data. Now, Glass and Smith and others have provided some widely accepted research findings relating class-size to both achievement (Glass and Smith, 1979) and affective outcomes (Smith and Glass, 1979). The key, it would seem would be to identify those policy areas that already are issues and try and provide as much evaluation evidence as possible even if the evidence cannot be collected by the state department.

The job description for my position has undergone considerable revision from that given on page 14 and puts emphasis on evaluation. as noted earlier, however, this titular emphasis has not been backed up with a staff adequate to the job responsibilities.

Evaluation related information has come into prominence in recent years--even if it has not always been used properly. A decade ago, no one in Virginia was particularly concerned about assessment. Now tests are everywhere. A thorough history of how tests use grew in kudzu-fashion is beyond the scope of this paper, but such a history might provide some clues about how to build interest both in other means of assessing outcomes and assessing other outcomes.

Finally, those interested in evaluation must just keep pushing evaluation as an important and useful activity. In the two and one half years that I have been with the department, workshops have been held both for persons in the field and in the department on techniques of evaluating projects and proposals. Each time a new program activity is proposed, the question is asked (by me) how are you going to evaluate it? The extent to which this awareness building activity has been productive is not fully known but there are encouraging signs that evaluation is being considered more as an integral part of programs from their inception. This is particularly true in formative evaluation where the information is used more as a guideline for further action than as summative judgment and hence maximizes the risk avoidance necessary in a bureaucracy.

In conclusion, it should be noted that "evaluation" is also changing. A decade ago, evaluation was conceived largely in terms of the laboratory model of research described by Fox (1977). The inadequacy of this model has been widely recognized and there are now a plethora of papers drawing from fields other than psychology and education to prescribe techniques for educational and psychological evaluation. There has been, similarly, a recognition by many that evaluation does occur in an environment where power, prestige, and economics often have a higher value than "scientific rationality". The degree to which changes in evaluation will produce models more relevant to more audiences is not clear. The degree to which evaluators will be willing to participate in "impure" research, the point at which they will feel that their hands are too politically dirty is also unclear. One must hope that the activities of evaluators will produce a better match between policy and results than now exists, and work to that end.

References

Alkin, Marvin C. and Daillak, Richard H. A study of evaluation utilization. Educational Evaluation and Policy Analysis, 1979, 1, No. 4, 41-49.

Astin, Alexander and Panos, Robert. The evaluation of educational programs. In Thorndike, Robert L., Educational Measurement, 2nd Edition. Washington, D.C.: American Council on Education, 1971, 733-751.

Berlak, Harold. Values, goals, public policy and educational evaluation. Review of Educational Research, 1970, 40, 261-278

Fox, G. Thomas. Limitations of a standard perspective on program evaluation: The example of ten years of Teacher Corps evaluations. Madison, Wisconsin, 1977.

Glass, Gene V. Standards and criteria. Presented at the 7th Annual Conference on Large Scale Assessment, Boulder, Colorado, 1977.

Glass, Gene V and Smith, Mary L. Meta-analysis of research on class size and achievement. Educational Evaluation and Policy Analysis, 1979, 1, No. 1, 2-16.

House, Ernest R. The objectivity, fairness, and justice of federal evaluation policy as reflected in the follow-through evaluation. Educational Evaluation and Policy Analysis, 1979, 1, No. 1, 28-42.

Jaeger, Richard N. A proposal for setting a standard on the North Carolina High School Competency Test. Presented to the North Carolina Association for Research in Education, 1978.

Mann, Dale. Policy Decision-Making in Education. New York: Teachers College Press, 1975.

McClung, Merle S. Are competency testing programs fair? Legal? Phi Delta Kappan, February, 1978.

Parlett, Malcolm and Hamilton, David. Evaluation as illumination: A new approach to the study of innovatory programs. In Glass, Gene V, Evaluation Studies Review Annual, Vol. 1, Beverly Hills: Sage Publications, 1976, 140-157.

Pipho, Chris. Presentation made at the 8th Annual Conference on Large-Scale Assessment, Boulder, Colorado, 1978.

Rich, Robert F. Translating Evaluation Into Policy. Beverly Hills, CA: Sage Publications, 1979.

Smith, Mary L. and Glass, Gene V. Relationship of class-size to classroom processes, teacher satisfaction and pupil affect: A meta-analysis. San Francisco, Far West Regional Laboratory for Research and Development, 1979.

Webster, William and Stufflebeam, Daniel. The state of theory and practice in educational evaluation in large urban school district. Presented to the annual meeting of the American Educational Research Association, Toronto, Canada, 1978.

THE INTERACTION BETWEEN POLICY AND EVALUATION WITHIN THE MICHIGAN DEPARTMENT OF EDUCATION

David L. Donovan
and
Stanley A. Rumbaugh
Michigan Department of Education

HISTORICAL BACKGROUND

The State Board of Education

The last fifteen years have been a period of transition in Michigan education. Historically the governance of education was delegated to local boards of education. The State Constitution, statutes and regulations tended to provide wide parameters for local programs and policy making. The Department of Education initiated little policy and was often criticized as a "do nothing" agency.

The early 1960s set the stage for change. Prior to 1964 the State Board of Education consisted of three members, plus the Superintendent of Public Instruction. The members, including the Superintendent, were elected at a biennial spring election. Since voter turnout at the spring election was usually small, educator organizations found it easy to influence the election of persons friendly to the concept of local control of education. The constitutional authority of the Superintendent was stated as "...shall have general supervision of public instruction in the state..." and, "...duties and compensation shall be prescribed by law."[1] The explicit authority of the Board was, "...shall have general supervision of the state normal college and state normal schools, and the duties shall be prescribed by law."[2] The State Superintendent and Board gave direction to a relatively small Department of about one hundred thirty professional and clerical employees. The authority was weak and the resources needed to govern a state educational system of over 700 school districts providing instruction to nearly two million students were inadequate. Thus, few policy initiatives emanated from the state to give direction to Michigan education.

The mid-sixties brought together several changes in thought, and several events in Michigan and the nation, to produce a different State Board and Department. The base change was incorporated in a new State Constitution adopted in 1964. It redefined the role of the State Board of Education, and changed the election process.

The 1964 Constitution established an eight-member board with candidates nominated at party conventions and elected at the regular fall biennial elections. Terms of office were made eight years. The membership of the Board was completely changed. A more subtle change than the election process and membership was the attitudes of the new members. They tended to have constituencies beyond education, to have social concerns beyond education, and to have political ambitions beyond the State Board of Education.

The new Constitution expanded the role and authority of the State Board of Education. The State Board duties were to be, "leadership and general supervision over all public education, including adult education and instructional programs in state institutions, except as to institutions of higher education granting baccalaureate degrees, is vested in a state board of education. It shall serve as the general planning and coordinating body for all public education, including higher education, and shall advise the legislature as to the financial requirements in connection therewith."[3] The persuasion of the Board was to develop a program of overall supervision of education and to initiate policies in keeping with their leadership mandate. The State Board was the policy board for the Department of Education. The State Superintendent was both the Chairperson of the Board and the Chief Administrative Officer of the Department. The Department of the past had to be changed to be responsive to the active role the Board wanted, and to the new world of education.

The Department of Education

The events of the time provided a setting. A collective bargaining statute was enacted, and old power structures were being altered. The courts increasingly entered into educational matters. Educational issues were often being identified and defined outside the educational community; there was a need for strong leadership in Michigan education.

The passage of the Elementary and Secondary Education Act in 1965 (ESEA) not only increased the Federal presence in education, it increased the role of state agencies in education. Most of the ESEA programs flowed program money through the state and provided administrative funds to the state. The resources available to the State Board and Superintendent were increasing and with it their ability to initiate policy was increasing.

As State Board members began to address their new responsibilities an obvious question was posed, how "healthy" is Michigan education? Answering the question was more difficult than expected. Although the Department collected some financial and staffing data and issued a few statistical reports, virtually no evaluation activities existed. No effort was made to gather and analyze a broad range of information about the schools and districts of the state. Certainly there was no effort to identify inadequacies, inefficiencies and inequities in the system. This paucity of information presented the policy makers a dilemma, a desire to provide the leadership for educational improvement, but no base of information about what changes were needed most. There were beliefs that some districts supported education at a much higher level than others, that learning levels were disparate, and that the conditions for educating were vastly different throughout the state.

Members of the State Board were anxious to perform in their leadership role and were not willing to wait until the Department could develop an evaluation capability and produce the information they needed. Rather, the State Board, with support from the Legislature, contracted with Dr. J. Allan Thomas of the University of Chicago to do a thorough study of Michigan education. His charge was to gather together information on the system, to describe the system, and to offer recommendations for improvement. The study took eighteen months and culminated with a report issued in the fall, 1967. The report drew several conclusions, among them:

1. There is a great variation in the educational opportunities available to students in the State of Michigan,[4] and

2. The Michigan State Department of Education should expand and strengthen the Bureau of Research, Planning and Development.[5]

The report was well accepted and thoroughly read by those interested in Michigan education. The report was a good base from which to set a direction. The goals of the 1970s were to be greater equity and equality in Michigan education, and evaluation, in the broadest sense, was to provide the leverage for the changes.

Initiating State Assessment

Acting on the recommendations of the "Thomas Study," the State Superintendent, Dr. Ira Polley reorganized and enhanced the evaluation capabilities of the Department by creating the Bureau of Research within the Department. Staff for the new Bureau was hired from bright, recent Ph.D. graduates of universities like Chicago, Columbia, Illinois, Oregon and Michigan. These persons not only brought new and different skills to the Department, but also the commitment to use these skills for educational improvement. Nearly as soon as the small staff of four or five were hired, they began to discuss the lack of reliable information about the status and progress of educational achievement in Michigan. This small group was familiar with, and intrigued by the research and writings of such men as Benson, Fox, Holland, Coleman, Thomas, Levin, Bowles and others.[6] The group generally embraced the input-process-output research model used by many of these investigators and saw as important to the state agency the answer to the question: "what are the correlates of educational success?"[7] Thus, staff discussions led to the development of a paper which suggested a statewide assessment effort to determine the status and progress of basic skills achievement and factors related to it. The paper was shared with the State Superintendent who was very positive and asked for alternative strategies for implementing the idea.

As the Superintendent and staff were considering the statewide assessment of achievement, the State Board was looking for ways to get information on the quality of education in Michigan. They were interested in the accreditation function then being fulfilled by the University of Michigan.[8] At its meeting of January 15, 1969 the Board discussed the accreditation process. It was obvious from information which came to the Board that taking over accreditation could not be easily accomplished...the University wanted to keep it. More importantly, the lack of any demonstrated relationship between accreditation and achievement was a major concern. Thus, when Polley introduced the assessment idea as an alternative, it was well accepted. Staff were asked to provide plans for a statewide assessment for the State Board's review. Proposals were placed before the Board in January,[9] February,[10] and April[11] and were thoroughly discussed and revised. In April the Board passed a resolution which directed the State Superintendent to seek legislation that would provide the authority and funding needed to carry out an assessment in 1969-70, and to do long range planning for a more comprehensive program. The Board emphasized that the basic skills assessment should also include information about the conditions under which the schools operated.

The legislature during the session had three other evaluation, assessment, or statewide testing bills introduced. In addition, the Governor's "Blue Ribbon" panel was about to recommend some kind of state assessment as part of the education reform package. While the task of getting the authority and funding for the assessment was not easy, the timing was right for approval. The State Superintendent was successful in gaining legislative support, and the assessment program was added to the Department appropriation bill for 1969.[12]

The Governor signed the bill in August, 1969. The mandate was to administer a statewide assessment of the basic skills prior to January 31, 1970. Staff began immediately to plan for an assessment which would yield reliable data on reading, English usage, and mathematics skill levels for Michigan school districts...provide an indicator of the level of basic skills achievement among the districts so that the disparities could be described and policies considered to address the problem areas. For the first time the state agency would be gathering information about the levels of achievement in the school districts of the state. The State Board would have information about the system it was to supervise.

Implementing State Assessment

The State Board in cooperation with the Legislature and Governor, had taken a major step by obtaining approval for a state assessment of the basic skills. However, before the first assessment was done in January, 1970, there was a changing of the guard. Ira Polley resigned as State Superintendent and was replaced by John W. Porter. In making this choice the Board made a commitment to a pro-active and highly visable role for evaluation. Porter brought a philosophical commitment to use of data in the management of the educational enterprise at both the state and local levels. To Porter evaluation was critical to managers. He was to define educational evaluation as, "a process of obtaining, for decision making purposes, information concerning educational activities,"[13] and emphasized his commitment by saying, "...we are committed to developing educational evaluation into a fruitful and productive exercise. We, in Michigan, are not content to treat evaluation as that useless exercise required from on high that takes time and pain to produce but which has very little significance for action." Porter, as State Superintendent during the 1970s, was to be the driving force behind state efforts in evaluation, and personally used the data provided to him.

Porter took office in October, 1969. The first state assessment was conducted the following January, 1970. The 1970 administration included the collection of data on student achievement as previously noted, but also included data on the socio-economic levels of the schools and district and general pupil attitudes. This was accomplished by administering a "General Information" questionnaire which contained twenty-six questions. Students responded to the questions anonymously. The purpose of the questionnaire was to provide the information needed to estimate the group socio-economic status, and the pupil attitudes toward self and school for each school and district. This was seen as necessary information to describe the conditions of education for the State Board, and for comparing group test results from year to year.

Some groups saw the questions in the "General Information" part of the assessment as unrelated to the purpose of assessment of the basic skills, and even worse, an invasion of personal privacy because of the questions asked as proxies for socio-economic status. The press picked up the complaints of educators and parents, and then legislators got into the debate. Department staff spent considerable time and effort explaining the need for these data, and defending its collection. Finally, as time passed and other issues arose, the controversy abated; but was to rearise each year until the State Board in 1973 directed the State Superintendent to eliminate the socio-economic status feature. It was recognized that these data were valuable for the proper analysis of the basic skills assessment data, but it was just not politically viable to keep this instrument as part of the program. The policy decision to eliminate it was made on political rather than technical grounds. At the same time the SES feature was eliminated, the questions used for constructing the attitude scales were eliminated because of technical deficiencies. Although these were corrected a year later, 1974, they were never reintroduced to the program.

Another controversy the first year was raised by legislators at the request of their constituents. They attacked one of the reading passages in the test because it was, "a blatant attempt to inculcate anti-American and anti-free enterprise values in school children."[14] The Department staff discussed these issues with the legislators and were able to avoid serious action against the assessment program. The compromise solution included changing the reading passage for the next year.

The time between passage of the assessment legislation and its first administration was only four months. The short timeline did not allow the development of the long range plan for assessment. The lack of a plan produced uncertainty and distrust among local educators over the ultimate purposes and uses of the tests and the data they would yield. This opposition was further stimulated because the program was new, it was championed by a new State Superintendent with whom they had little prior experience, and it was an intrusion by the state into local educational autonomy. The State Board and Department were seen as pushing out to exercise new authority and were using the assessment program as a vehicle. The program became the focal point of opposition.

The results of the first assessment were sent to local school districts without fanfare. There was some interest from the press in reporting "scores" of local districts, but the Department deferred requests to local districts, or were able to convince the media people not to report them.

The Department used the results for two major purposes, 1) analysis of the correlates of educational success, and 2) as one of the criteria to determine school district eligibility for state compensatory education funds. The analyses to identify correlates of school success confirmed other such studies; the correlations between mean achievement as measured by the test and percent minority students were about .50[15] and achievement and mean socio-economic status of students were about .60.[16] The analyses were disappointing because all other correlations between composite achievement; and expenditures, staff training, salaries; size of district, etc. were less than .20.[17] Again the variables in control of the school managers were disappointing. The Department was to repeat the studies the next year with the same results, and then dropped such analyses from the program.

The second major state use of the results was as one of the criteria for determining school district eligibility for funds under the state compensatory education program. Eligibility for funds had previously been determined on the basis of socio-economic indicators (e.g., similar to the ESEA Title I use of aid to dependent children, family income, etc. as indicators). There was a strong feeling among Department staff led by John Porter that eligibility should be more directly determined by a measure of "educational deprivation," i.e., low basic skills achievement. A position paper which advocated the use of mean district scores on the state assessment as one of the criteria for eligibility was developed, and adopted by the State Board of Education. The idea was well received in the legislature, and the state compensatory education program (Chapter 3 of the State School Aid Act) was amended for 1971. The policy makers strongly believed the direct achievement measure was a better criterion for directing funds to alleviate low achievement problems. Later, Chapter 3 was amended to make results on the state assessment the sole criterion for district eligibility. Chapter 3 will be more thoroughly discussed in a later section of this paper.

State Assessment: Local Educator Reaction

The state assessment the first year had been authorized and funded through the Department budget bill. This was an expedient method, but only a temporary one. The "Governor's Task Force on Educational Reform" had reported the need for a continuing measure of pupil achievement. The Department staff developed a draft bill which provided continuing authority for the program. Staff were concurrently working on revisions to the state compensatory education legislation to include educational deprivation in the criteria. It was natural that the draft bill to authorize state assessment as a permanent program would tie-in compensatory programs. The draft bill with some minor changes was enacted as Act 38 of the P.A. of 1970, and remains as the legal base for the program today. The legislation broadly states: "A statewide program of assessment of educational progress and remedial assistance in the basic skills of students in reading, mathematics, language arts and/or other general subject areas is established in the department of education..."[18]

The provisions then go on to give various elements of the program. Included are: establish achievement goals, provide information useful for allocation of state funds to equalize education opportunity, provide incentives to introduce programs to improve basic skills or attainments, and provide the public information on the school system. With this legislation, the state assessment had a definite mandate.

The assessment tests from 1970 were revised for use in the 1971 assessment. The tests were lengthened so that each fourth and seventh grade pupil would take a full test battery for reading, mechanics of written English and mathematics. The tests would now yield scores reliable enough for reporting individual pupil socres, as well as aggregate scores at the school building and district levels, and a state report. The tests were administered throughout the state in January, 1971, along with the still controversial socio-economic status and attitude scales.

Before the test administration period was over, a group of local superintendents met to review this "new" state program of assessment. These discussions led to action by some thirty-eight of them. They ordered that the test answer sheets be held in the district and not sent to the scoring service. The press picked up the story and the state assessment became a big story...the program had visibility!

After two weeks of unsuccessful discussions where state officials tried to convince the superintendents to send in the answer sheets for scoring, the State Superintendent and the President of the State Board of Education sent a joint letter to local superintendents and board presidents.[19] The letter cited "Act 38" authority for the assessments, directed the submission of answer sheets, threatened court action and offered to discuss the superintendents' concerns. The superintendents, though reluctant to comply, chose not to challenge the state authority further.

In the ensuing discussions local superintendents raised several issues. The major issue was, of course, the intrusion of the state into local school affairs. Each of the seven or eight meetings between Department staff and superintendents began with this issue and required a rejustification of state assessment and the state authority. Other charges were made, such as, 1) the tests were invalid and did not correspond to the "Michigan Curriculum," much less the "unique" curricula of the many local districts; 2) the tests were ill constructed and unreliable; 3) the test information was no different than that already known from local testing; 4) local educators had no opportunity to participate in the planning of the state assessment program; and 5) reports of results would be made to the public. The criticisms were in part acknowledged by the staff and promises were made to be more responsive to the involvement and data needs of local educators. The technical issues, i.e., reliability and validity of the tests, were defended and final disposition of the charges were left to the publication of the technical reports.

After nearly seven months of monthly meetings, the superintendents, though still not satisfied, decided further discussions were unnecessary. They would cooperate in the future, and the Department would form an advisory council to help form the future of state assessment.

Public Release of Results

The promise had been made by Department officials in the first year of the program that the state would not release the "scores" of individual school districts. With this promise, local school officials felt secure with the new program and cooperated in the administration with a minimum of objection. Scores were reported directly to local school superintendents and to state officials.

In the second year, the press, and eventually legislators, made inquiry about the "scores" and were told they could not be released. This led to a confrontation. An influential legislator threatened to introduce legislation which would mandate the release of the state assessment data for schools and districts and would provide guidelines for the release. After discussions with the legislator, the Deaprtment policy on release of these data was changed and the basic skills assessment data were provided to the press and legislators. Even today, ten years later, the promise which couldn't be delivered, i.e., no public release of school or district results, is remembered by some superintendents.

The first release of results was made in response to individual requests. However, the interest was great and the State Superintendent decided to publish the results for all districts. The first to be published were the 1971 results. A compilation of data (assessment test results, staffing, financial, dropout, etc.) was made. A book was put together and released about a year after the test administration; ironically the book had a red cover and the press conference for its release was on Valentine's Day, 1972.[20] Superintendents had no love of the assessment, and saw no humor in this. The 1972 results were published in like form...the book being brown and the release was at Thanksgiving, 1972.[21] The "thanks" of local school people was that this was the last book of all district results published by the state. The "heat" was too much and the Superintendent decided results from 1973 were to be released on request, but no compilation of all districts was released.

After 1973, rank ordering of school districts was not done. The tests were changed from norm referenced, which could be reported succinctly in standard scores and percentiles, to objective referenced, which were reported in proportion of pupils mastering each objective. Since there were over sixty objectives, the publication of even district level data for all districts was too burdensome. Assessment results were made available on request, and often were listed in the newspapers.

The public release of results was made even more necessary in 1977 when the State of Michigan enacted the Freedom of Information Act which required public agencies to make information available upon request.

The policies on release of assessment scores have been influenced by educators, legislators, public advocacy groups, and assessment technicians. The "pull and tug" to derive a policy involved the desire for widespread public disclosure on the one hand, and the fear of misinterpretation and misuse on the other. School people feared invidious comparisons of schools, and judgements of school effectiveness based on a narrow set of measures. Public advocacy groups pressured for full disclose and specialized reporting for subpopulations, e.g., racial-ethnic groups. Assessment technicians counseled caution in generalizing from the data and sought ways to provide better interpretive reports. Legislators, at first, wanted full disclosure, but have more recently pressed for recognition of the limitations on the data. In fact the Legislature inserted a statement into the Department budget bill which prohibits the use of assessment as an evaluation of schools. The current policy of disclosure is to make results for a school or district available upon request, but to provide explanatory and interpretive materials along with it. The State Board in 1979 sponsored several workshops for local educators and the press. The Board's purpose was to assist local educators to work with the press to achieve full disclose with responsible reporting. Also, the State Board adopted a policy which stated that assessment results were not appropriate to use in the evaluation of an individual teacher.

Evaluation of Schools

Tied closely as an issue to the public reporting issue are the issues of use of assessment results in making decisions. How good is my school or district has always been the prime question of interest to citizens. Before state assessment the judgments undoubtedly were made on criteria ranging from hearsay, to athletic teams, to the number of graduates getting scholarships, to any of numerous other factors. State assessment was of interest as a criterion for judging the worth of a school or district because it was reporting on how well pupils were learning. After all, schools existed to teach the basic skills and should be rated on how well this was accomplished. Newspapers reported scores and pointed to high scoring districts (mean of pupil scores in the district was used, and later the proportion of pupils who mastered more than 75% of the objectives tested) as "good," and the low scoring as "poor." Real estate agents, too, tried to use the scores, if it suited their purpose, to steer customers to buy in "good" districts.

The comparison of school and districts on assessment scores alone was a concern for school administrators. They carried their dissatisfaction to some key legislators as well as State Board members. Under pressure from the legislators, the Department initiated a large campaign to assist local educators in the proper and full reporting of results. Advocated were early reporting, and reporting in the context of other information about education, i.e., the financial, staffing and other conditions of education. The idea was to put the assessment scores in a larger context to provide for a fuller understanding and a better "evaluation" of the schools than a simple judgment made on one set of test scores.

Maximize Policy Use or Instructional Use

Initially (i.e., 1970-73) the state assessment program used norm referenced tests developed by a testing company from existing items to the Department specifications. The primary purpose of the program was to measure the status and progress of basic skills achievement in the state and its districts. These tests in reading, mathematics and mechanics of written English provided data for these purposes with a minimum of expense and testing time. An aggregate achievement "score" for the state in each area was computed, as was a "score" for each district in the state. Districts were easily ranked by percentiles, and districts in need of assistance were easily identified. State policy purposes were well served by the use of norm referenced tests.

Politically though there was discontent with the program. The discontent involved: 1) the use of the results to compare school districts, 2) the tests did not provide information useful to schools in instruction, and 3) the tests were not "Michigan tests" and Michigan educators had not been involved in creating them.

After the furor created by the superintendents in 1971, the State Superintendent decided to both be responsive to the issues raised about the norm tests, and to exercise state leadership in basic skills curricula for the state. It was decided to change from norm referenced to objective referenced tests for the state assessments. The decision would switch the emphasis to maximize the instructional and curriculum uses of the results, at the local level, rather than the policy use at the state level.

The State Superintendent met with each of the statewide curriculum organizations (i.e., mathematics, reading, science, social studies, health education, physical education, art, music) and challenged them to specify the basic expectations for their area. The basic expectations were, in general, defined as what every pupil should be able to do and should know at the end of grade 3, grade 6 and grade 9. These were to be "minimal expectations" for all pupils in Michigan schools and would be strongly advocated as the minimum curricula for all schools in the State. The curriculum organizations, after much discussion, all chose to respond and work with Department curriculum specialists to specify the "minimal expectations."

During late 1971 and early 1972 the curriculum specialists drafted the expectations. These were reviewed and, in some cases, revised by committees of generalists (i.e., teachers, principals, school board members, school administrators and parents). Finally, in 1972 the State Board adopted the first two sets of expectations, or objectives. These were the reading and mathematics objectives which were to be used in the new state assessments.

The tests were constructed from the objectives. The Department engaged some local school districts to provide: 1) teachers to write test items, and 2) classrooms for test tryouts. The tests were to be written primarily by Michigan teachers based on Michigan produced objectives...these were to be "Michigan tests." The Department also contracted with a testing company for support services to insure that the new tests would meet high technical standards.

The tests were completed and ready for use in the fall, 1973. The test administration time was changed from January to September-October with the initiation of the objective referenced program. This was done because of the emphasis on instructional uses. The early administration allowed the return of results early so that individual pupil needs could be identified and teachers would have time during the school year to provide remediation, if needed. The reports would identify the objectives mastered, and those not mastered. Mastery was defined as answering correctly at least four of the five test items for each of the forty mathematics and twenty-three of the reading objectives. The reports contained detailed information compared to the general information contained in the norm test reports. The detail of many scores made it more difficult to compare schools and districts on the basis of state assessment, but made the information more valuable to principals and teachers.

The State Board had used the state assessment program to exercise leadership in Michigan education. For the first time a common curriculum had been specified, albeit it was only a minimal level and was suggested rather than mandated. The minimum expectations though were to become useful in promoting equal educational opportunity initiatives in schools of the state.

State Level Uses

The change to objective referenced tests and the more detailed reports was responsive to local education criticisms. However, when the first reports were released, the press, and state officials were confused by the many figures. They wanted to be able to tell whether or not schools were doing better than last year, and which were "good" and which were "poor" achieving schools. There was a demand for a simple summary type report. The State Superintendent asked for a single score.

The political pressure from State officials led to the development of a summary type report. The report was added in 1974 and was called the "proportions report." The report gave the percent of pupils mastering objectives in each of four categories (i.e., 0-24, 25-49, 50-74, 75-100 percent). The reports were in reading and mathematics and were produced for schools, districts and the state. The fewer figures were more understandable and useful to laypersons and for state purposes.

The proportions reports were used to set criteria for identifying levels of needs in Michigan schools (e.g., schools with fewer than

50 percent of the pupils mastering 75 percent or more of the objectives were defined as high needs schools). The State Superintendent and staff directed special assistance to these schools in an effort to assist them to improve.

The assessment program reflected an action policy of the Department of Education to seek and use information to develop a concerted program for educational improvement. However, it was too general to provide information to assist in determining the success or effectiveness of specific program efforts. Thus, concurrent with the development of the assessment program, the evaluation program was also being developed.

Initiating Program Evaluation

As noted in an earlier section, the advent of extensive federal involvement in education, in particular, the Elementary and Secondary Education Act of 1965 (ESEA), provided a new impetus for State Education Agencies across the nation. In addition to creating an expanded role for State Education Agencies in educational program development and administration, ESEA demanded a more active role for State Education Agencies in evaluating those programs.

ESEA caught most state and local education agencies unprepared to undertake sophisticated and technically sound program evaluation. Federal officials were vague in providing direction and frequently suggested summative questions which needed response. These questions were usually descriptive as well as summative in nature. Beyond these basic descriptive questions, State Education Agencies were encouraged to develop evaluation capabilities and design evaluations to best meet the needs of state and local constituents.

In Michigan, as in most states, early evaluation efforts were aimed at meeting the summative evaluation requirements. The prevailing philosophy was that evaluation was a federal reporting requirement which had to be done in order to maintain eligibility for funds. These "reporting" activities were decentralized in the state agency as part of the overall responsibility of the persons who administered the programs. The evaluation results were seldom used (nor thought to be useful) in program administration or policy development.

The decentralized approach and the "required reporting" philosophy toward evaluation began to change in Michigan in 1969 with the creation of a Bureau of Research. With the establishment of a new Bureau, came the direction for the new staff to begin conducting evaluation of the new federal programs, and to use some of the federal money to support these evaluations. This new commitment was further strengthened by the appointment of a new Superintendent who, as noted earlier, believed that information provided by technically sound evaluations would lead to improved decisions regarding educational programs.

The early and active support of the State Superintendent resulted in a decision to begin recruitment and employment of a small number of specialized staff and to begin centralizing the function of evaluation. Evaluation staff were to be administratively independent of the personnel who were responsible for management of the programs to be evaluated. The new staff were asked to develop and implement a systematic approach to program evaluation.

In 1974, the State Superintendent emphasized his support of evaluation and put the full weight of his office behind the centralization of the evaluation functions in the evaluation program, was issued. Portions of the policy statement follow.

> It is my intent that all evaluation activities sponsored by the Department of Education be coordinated by staff with expertise in evaluation so as to maintain consistency in the evaluation efforts.[22]

After citing several negative aspects associated with decentralized evaluation efforts, the Superintendent's statement continued:

> If evaluation is worth doing, it is worth doing well. Furthermore, program administrators should never evaluate their own programs. Therefore, effective immediately, I am asking each of you to ensure that the evaluations of your programs are coordinated through (the Evaluation Program) which is...responsible for evaluation.[23]

The statement concluded by indicating actions which should be taken to receive approval for evaluation activities.

In 1977, the State Superintendent repeated his statement verbatim and added that he expected any items which included plans for evaluation, such as programmatic state plans, to include a statement of support from the evaluation staff before being submitted to the State Board of Education for approval.[24]

The Evaluation Process

The goal of evaluation was to provide information to educational decision-makers so program improvements could be made. Staff were committed to the task of demonstrating that well-designed, carefully implemented and properly supported evaluation "provides objective information for planning, administering and improving educational services at all levels of educational governance, from federal and state to school district, to school building and to classroom levels."[25]

In an enterprise so large and encompassing as education with so many factors beyond the control of the evaluation specialist, it is impossible to employ the same experimental rigor which might be found in a scientific laboratory. In the social sciences and education it is often impossible to control conditions and set up experimental designs as in the natural sciences Therefore, the evaluation model employed by the Department had three stages: (1) descriptive evaluation, (2) evaluation to determine success, and (3) evaluation to determine effectiveness. Descriptive evaluation refers principally to the quantitative description of resources (human, financial and material) and purposes associated with educational services. Evaluation of success refers to quantitative and qualitative judgments regarding whether or not objectives of an educational delivery system have been met. Evaluation of effectiveness refers primarily to identification of factors associated with success and the relative costs of assuring that those factors exist.

While these stages are sequential in nature, they are also fluid and overlapping. For example, it will take some evaluations a year or more to pass through the descriptive evaluation stage while others will pass through this stage much more quickly. Also, work may be occurring in more than one stage simultaneously, for example, evaluation of success may begin before the descriptive evaluation is complete. Furthermore, the implementation of each successive stage does not mean that the prior stages are terminated. Rather, each successive stage builds upon the information provided by the preceding stages.

In the last half of the 1970s several evaluations were able to identify factors which are related to success. Based, in large part, on these evaluations the Department is now exploring means by which, through a state-local partnership, strategies can be developed which will lead to more predictable program improvement.

As part of the State Superintendent's policy statement,[26, 27] he asked program administrators to enter into "Agreements for Services" with the evaluation staff for the conduct of evaluations. The "Agreements for Services" reached between program administration staff and evaluation staff specify services and responsibilities of both staff and formalize expectations of both parties. The agreement commits the administrative unit to provide a mutually agreed amount of funds to conduct the evaluation. The agreement commits the evaluation unit to provide information to answer specified program and policy questions. Staff employed for the evaluation are administratively and programmatically independent of the program administration staff.

While administrative independence is desirable, daily substantive interaction among the evaluation and program administration staff is essential. Evaluators must be aware of, and sensitive to, the subtleties of the program they are evaluating. Also, informal substantive contact decreases the threat often associated with evaluation.

Much emphasis is placed on communication among staff and the appropriate use of evaluation results. In addition to the day to day contact among staff, more formal mechanisms are used for presenting findings and recommendations. Formal "exit conferences" are attended by evaluation staff, program administration staff, and, frequently, one or more high level officials of the Department. At these "exit conferences," findings of the evaluation and action-oriented recommendations are formally presented by evaluation staff. Program administration staff respond, either at the "exit conference," or soon after, regarding actions they plan to take on each recommendation. The "exit conference" reduces the likelihood that evaluation findings and recommendations will be ignored. The administrator in charge of both the program and evaluation responsibilities is present, supportive and can direct actions and policy responsive to the evaluation findings.

The State Board of Education has, historically, been very interested in the work of the evaluation staff. Great care is taken to prepare and present evaluation reports to the State Board which will be meaningful as a tool to guide in establishing policy. Frequently, major segments of time are set aside by the State Board of Education at Committee of the Whole meetings to discuss evaluation reports, recommendations and implications for administrative and State Board action.

In addition to formal and informal communication efforts with Department staff and the State Board of Education, evaluation staff are actively involved in a program of technical assistance and dissemination to local education agencies. These activities cross a broad range from distribution of executive summaries of evaluation reports to formal inservice or technical training sessions.

The Conflict Between Expectations and Methods

There has been a strong policy and programmatic commitment to evaluation in the Michigan Department. An environment has been created to promote the development of a strong organization for conducting evaluation. The commitment to use evaluation findings and act on evaluation recommendations has been high on the part of the staff of the Department and the State Board of Education. But even in an organization with this high level of commitment conflicts between expectations and methods can and do occur.

One of the most common areas of conflict is a conflict among program priorities. These conflicts are primarily of two types. One type of conflict is a result of a lack of clear enunication of the purposes and priorities of a given program. Evaluation staff, during the "descriptive stage" of evaluation, work with program administrative staff to clarify the purposes and objectives of the program to be evaluated. This process can be a particularly frustrating one, if there are external pressures to provide information quickly. The lack of clear enunciation of purposes and priorities becomes acute when the evaluation effort begins a lengthy period of time after the program begins.

The examination of program purposes and priorities frequently leads to a second type of conflict among program priorities. This occurs when the programs have a mixture of social action and education priorities. For many reasons, categorically funded programs often have a multiplicity of apparent purposes; some establish primarily education priorities while others establish primarily action priorities. For example, legislation may contain language which seems to equate civil rights and basic skills education.

It is not uncommon for these social action and education priorities to be so closely intertwined that it becomes virtually impossible to distinguish among them. The "descriptive" stage of evaluation is used to deal with this problem (as a part of enunciation of purposes and priorities). However, even if the social action and education priorities can be identified and separated, these programs are especially difficult to evaluate. In some cases, program administrative staff have preconceived expectations regarding the outcomes of an evaluation. Additionally, they often do not fully understand that social action objectives cannot be measured by educational performance measures. This combination of preconceptions and misunderstandings can lead to great disappointment upon the completion of the evaluation.

Evaluation must be especially careful to develop mutual understanding about purposes, priorities and expectations of the program to be evaluated. Of equal importance, is the development of mutual understanding regarding expectations for the evaluation effort. Evaluation staff of the Michigan Department of Education use the "descriptive" stage of evaluation to develop these understandings. However, the affirmation of these understandings must be continuous.

A second area of common conflict between expectations and methods is the conflict in requirements. This type of conflict is the conflict between program funding mechanisms and expected results, and most often occurs when programs are funded on one set of criteria and the program success is judged on a different set of criteria. For example, the funds are provided for reimbursement of program staff salaries, but the evaluation focus is on how much the participants achieve. This particular type of conflict may create hostility among local education agency staff who feel that it is unfair to conduct a state-level evaluation of those parts of the program funded locally.

This type of conflict often establishes a negative political environment within which it is very difficult to conduct an evaluation. Evaluations fraught with this type of conflict usually will not advance past the "evaluation of success stage." The evaluators can use the earlier stages of the evaluation to establish reasonable measurement criteria and data collection procedures. However, there are likely to be so many negative factors beyond the control of the evaluators, that some such evaluations will never leave the "descriptive evaluation" stage.

Another area of conflict occurs between federal constraints and state-local policy and program needs. Historically, the requirements of federal programs have focused on summary reporting and are of little use in state or local program or instructional decision making. This is not a problem so long as state and local education agencies are able to exceed these requirements. In fact, if the burden of federal reporting is minimal and federal funds can legitimately be used to expand the evaluation of that program to yield results meaningful to state and local educators, a positive state of affiars exists. The conflict occurs when federal requirements, even though summative in nature, are so burdensome that all of the resources available must be used in meeting the federal requirements. A review of the history of evaluation of ESEA Title I in Michigan suggests that this pattern has occurred. ESEA Title I will be discussed in more detail in a later section of this paper.

Recently, a second type of conflict between federal program requirements and state-local policy and program needs for evaluation has become common. Federal programs are becoming more and more prescriptive with regard to mandating specific information which must be gathered and the specific evaluation procedures which must be used by state and local evaluation staffs. These procedures are frequently insensitive to state and local policy and program needs. Further, the constraints are such that states can do little to design evaluations to meet both federal requirements and state and local needs. An example of this "federal prescription service" are the rules and regulations dealing with data collection and evaluation of programs funded under PL 94-482 (Vocational Education Amendments of 1976).

PL 94-482 and its associated rules and regulations require both evaluation and reporting of management data (the Vocational Education Data System--VEDS) on every vocational program. The evaluation requirements, by themselves, are manageable and considered by many to be useful. However, the reporting requirements of VEDS are so prescriptive and burdensome that the Council of Chief State School Officers (CCSSO) has officially opposed them and threatened to refuse to comply. This enormous data collection burden imposed by the federal government has made it very difficult for evaluators to collect and analyze new data needed for more meaningful evaluations.

A third general area of conflict is a conflict in commitment. Frequently, top level policy makers do not provide adequate support for evaluation activities because they have an incorrect impression of what evaluators do. This is especially true if the only visible product of the evaluation effort is an annual summary report which has little perceived usefulness. Evaluators need to do a much better job of helping policy makers understand what evaluators do. Formal and informal communications must not stop with program administrators.

Another problem is that commitment from top level policy makers, as evidenced by resource availability, is inconsistent. Ironically, in periods of economic difficulty, resources for evaluation may actually increase as decision makers seek data to help with the management of decline. In economically good times, evaluation may not seem as necessary and resources for evaluation may become less plentiful. This inconsistency, even in a state with a generally high level of commitment, makes long range planning for evaluation somewhat more difficult than desirable.

A Case of the Influence of Evaluation on Policy

Michigan has had a state compensatory education program since 1967. The first years of the program funds were distributed to school districts as formula grants. The formula used economic, cultural and social factors for defining educational need. However, beginning in 1970 Michigan began to define educational need in terms of pupil achievement, i.e., a direct measure of educational need rather than a proxy measure. The state compensatory education program makes use of the state educational assessment results for this purpose. The procedure for determining school district eligibility, beginning in 1971, for the compensatory funds was:

1. Pupils scoring below the 15th percentile on the 4th and 7th grade state assessment test were defined as pupils to be counted as "eligible."

2. The proportion of all 4th grade pupils deemed "eligible" was computed; likewise the proportion of the 7th grade pupils.

3. Applying the proportion of eligibles in grades 4 to grades K-3, an estimate of "eligibles" in those grades was computed; likewise the grade 7 proportion was used to estimate grade 5-6 "eligibles."

4. The school districts were ranked (high first) according to the proportion of "eligibles" in the district.

5. The district allocation was computed by multiplying the number of "eligibles" in grades K-6 (the State program was limited to these grades) times $200 (the funding level). Districts were funded in rank order until the total State appropriation was used.

It was determined that the State could afford $22.5 million for the program before the formula for the program was written into legislation. Legislators used several computer simulations, each with different eligibility and/or fund level criteria, in the process of setting criteria. Basically the data were used in making political-policy decisions. Legislators wanted to know which districts would be funded, or not funded, and at what level, before agreeing on the formula. The final formula was a compromise made by members of the appropriations and education committees of the legislature.

The use of an achievement test indicator for determining the level of educational need was but one of several different features of the Michigan program. Others were:

1. Assurance of three years of funding once a district was deemed eligible.

2. Provision for funding adjustments based on program success.

3. Provision for annual evaluation of each pupil's progress to determine level of attainment.

4. Provision of considerable local discretion vested in local districts in the use of funds.

Department staff worked together with local educators to design the program. The local educators were interested in three things in the new compensatory education program: 1) more money, 2) more discretion in the use of the money, and 3) greater assurance that the money would be available for more than one year. Each of these was attained in the new legislation.[28]

The State in this program was interested in two important propositions: 1) could schools be held accountable for educating the lowest achieving pupils in the schools, and 2) could additional money for basic skills instruction result in higher pupil achievement?

The program has changed over the years since 1971. When the State assessment changed to objective referenced tests in 1973, the criterion for eligibility was changed from students below the 15th percentile to students achieving fewer than 40% of the objectives tested. Again the legilsature used several simulations of data to set the formula so that the funded districts and the funding level remained comparable to the previous years.

The three years funding feature was later changed to an annual redetermination of district eligibility and funding level. Non-funded districts lobbied for the change so they would have a chance for funds before the three year cycle was completed.

A philosophical belief was that all pupils regardless of race, geographical location, economic status, etc. could attain basic reading and mathematics skills. Thus, a feature to reduce funding if pupils didn't achieve was included. The adjustments were to be made annually on the basis of pupil achievement. Pupils who achieved at least 7.5 months gain, as measured in grade equivalent units on a standardized test, received a full $200 allocation for the next year. However, if achievement was lower than 7.5 months, the district received a lesser amount (the proportion being the gain in months to 7.5 times $200). Local districts accepted this feature initially to get the money. After the first year they lobbied to retain the money which was going to be "lost" because "the kids still need the help." The money was reallocated to the district provided they filed a plan to meet the needs of the students who were still low achievers. After two years, and threats of losing more money, the districts succeeded in getting the legislature to delete this "accountability" feature from the program.

The evaluation of the compensatory education program was linked very closely with the "accountability" feature. Since the funding was determined on a per pupil basis and that was tied to level of attainment, it was necessary to evaluate on an individual pupil basis. State guidelines called for a pre and post test (either spring to spring, or fall to spring administration) using approved standardized tests. Scores for each pupil were submitted to the State and were used both for program evaluation and for the determination of funding. The verification and processing of over 112,000 pupil records was quite a challenge for State evaluators.

The evaluations of the program showed the program to be a success. The districts committed themselves to developing quality basic skills programs based on specific performance objectives. Strategies were developed to provide services to low achieving children regardless of the school attended, thus moving away from the "target" school concept. Most important of all, the program resulted in improved achievement for pupils in the program.

The State evaluators not only analyzed data from the school districts to determine program success, made recommendations for program improvement and provided funding allocations, they also used the program to improve evaluation techniques across the State. The evaluation of individual achievement, as well as program evaluation, presented many local school educators a challenge beyond their knowledge and skill level. State staff were able to seize this opportunity to provide inservice training to improve evaluation methodology and data use in many school districts. Particular emphasis was placed on working with local district staffs to develop objective-referenced tests for evaluation purposes.

Unfortunately, the elimination of the funding adjustments based on the success feature made some people believe there was little need to continue program evaluation. Thus, funding to maintain the State evaluation staff was deleted from the Department budget, even though for another three years the mandate to provide the legislature with an evaluation report remained in the act. Local schools, in most cases, continued the program evaluation and used the results locally, however, State activities stopped with the withdrawal of funding.

At the time evaluation funds were deleted, it was suggested that State assessment results be used to evaluate the program. The belief was, since State assessment was used to determine eligibility for compensatory education funds, the same test, over the same objectives, should be used for evaluation. Very simply, it was thought the fourth grade results would show success through the first four years, and seventh the last three years of the program. There was a certain logic to the proposal, however, there were many falacies: 1) State assessment results were reported for the total pupil population, not compensatory pupils as a subpopulation; 2) pupils moved in and out of compensatory programs and treatments varied; 3) one measure was not sufficient to evaluate program success and would tell nothing about why some programs succeeded more than others.

The proposal was hotly debated in the Department, and an attempt was made to implement this "evaluation" by identifying individual pupil assessment data with the compensatory services the pupil received. Local school administrators and evaluators strongly opposed such "evaluation."

The local staffs refused to cooperate in the coding and the attempt failed. After negotiations it was decided to use state assessment as a vehicle for collecting some data about the compensatory program. Local evaluators agreed to code pupils on the fall assessment as enrolled in the various compensatory programs, i.e., ESEA Title I, State, Bilingual. This allowed the State to address some questions with policy implications: 1) were the lowest achievers in compensatory programs, and 2) were pupils enrolled in more than one compensatory program. In case the lowest achievers were not in the program, more stringent guidelines for pupil selection could be imposed. Data for the second question would be used to address whether or not the greater benefit was to continue multiple funding or spread money to new pupils. The coding project is now in its first year and data are being analyzed. As for the more in depth questions of evaluation, local districts and the Department are cooperating in special case studies. These studies will address the reasons some programs are successful and others are not.

The State Compensatory Program is a good example of using data to carry out a philosophical belief, i.e., low achievers should be benefited by funds. Also, it was a good example of evaluation being used in the management of the program and its improvement. Unfortunately, the evaluation efforts were not appreciated and resources were withdrawn prior to the potential benefits being attained. The program continues, but there is no way of systematically judging its effectiveness.

Title I Evaluation - A Case of Categorical Constraints

The preceding discussion of Article 3 evaluation has presented a case study of evaluation's impact on policy. The evaluation of ESEA Title I is a case study of the categorical constraints on the usefulness of evaluation in policy development.

Title I evaluation development paralleled, in many respects, the evaluation of Article 3. In fact, this federal program, more than any other single program, provided impetus for evaluation in the Michigan Department of Education.

In the late 1960's and early 1970's, the federal reporting requirements for ESEA Title I were quite minimal. Summative information of a descriptive nature was expected and achievement data were desired by federal officials. However, it was recognized that many state and local education agencies did not have the capability to conduct more sophisticated evaluations. Federal officials encouraged State and local officials to develop capability to do evaluations that exceeded the minimal expectations.

Thus, during the late 1960's and early 1970's the evaluations of Title I conducted by the Michigan Department of Education were almost entirely descriptive, consisting initially of baseline information such as number of students, number of teachers, and amount of money spent. Beginning in 1972, the evaluation started to yield accurate and useful information regarding success (in terms of achievement of students) of the program across the State. This information was based on district level information. Through the 1974-75 school year the design remained relatively constant so as to verify findings of success.[29]

State and federal officials were able to say with considerable confidence that Title I in Michigan was successful. However, the "success evaluation" did not provide sufficient information to enable state and local officials to identify or select specific strategies associated with success. These might be used for improving local projects which were not successful. Consequently, since 1975-76, the evaluation of ESEA Title I in Michigan has focused on the building as the level for data collection and analysis.

In the evaluation of Title I, the Department of Education evaluation staff have been successful in identifying a number of variables related to success. Further, these variables have been verified by other studies. Thus, evaluation of ESEA Title I in Michigan has advanced into the "evaluation of effectiveness" stage discussed earlier in this paper.

In addition to the progress in increasing the sophistication of the state level evaluation, evaluation staff have worked with local education agencies to make local evaluation efforts more useful. For many of the same reasons the assessment program switched to objective referenced tests, local districts were encouraged to develop objectives and objective-referenced tests for evaluation purposes. Together, State and local officials worked to make the objective-referenced tests useful both for instruction and evaluation. In order to assure high standards of quality for locally developed objective-referenced tests, Department staff produced a quality-control system for objective-referenced tests[30] which was generally followed by local districts. Locally developed tests had to meet the standards of this quality control system before they were approved by the Department for use in evaluation of ESEA, Title I.

The development and planned use of objective-referenced tests for Title I evaluation peaked in 1974-75 and 1975-76 at about the same time that the state level evaluation was reaching the "evaluation of effectiveness" stage. Thus, the Michigan Department of Education was able to take advantage of the impetus provided by Title I to develop sound evaluation procedures which yielded meaningful results for policy development and program improvement at both the state and local levels. Additionally, the data provided to the U.S. Office of Education about Title I in Michigan was of high quality. However, not all states had developed a high degree of sophistication and those which had done so had used different methods and procedures. In short, the federal policy of encouraging development of evaluations which were useful at State and local levels had resulted in data at the national level which were not comparable and of varying degrees of quality.

In testimony during the debate leading to reauthorization of ESEA Title I in 1974, Congress expressed considerable dissatisfaction with the lack of comparable data to guide its deliberations. This dissatisfaction was specifically exhibited in the Education Amendments of 1974 and the Education Amendments of 1978.

The legislation required the U.S. Office of Education to develop and provide to state education agencies "models for evaluation of all programs"[31] funded under Title I "...to be utilized by local educational agencies, as well as by the state agency in the evaluation of such programs."[32] The law further stipulated that the models should yield data which are comparable on a state and national basis.[33]

At the time that U.S. Office of Education began initial development of these evaluation models, Michigan Department of Education staff were struggling to develop procedures for aggregating data from locally developed objective-referenced tests. It was hoped that the new models would recognize the value of objective-referenced tests and that a sound model for their use would be developed. It was soon learned, however, that such was not to be the case.

It became obvious that the models being developed would be of limited usefulness to the Michigan Department of Education. Consequently, while supporting the need for nationally comparable data, the evaluation staff of the Department actively advocated the development of more flexible models aimed at identifying variables associated with achievement and greater utility at the state and local levels.[34]

In October, 1979, the final rules and regulations[35] were passed mandating the use of three evaluation models. The models are much more restrictive and less useful at the state and local levels than hoped for by evaluation staff of the Michigan Department of Education.

The immediate effect has been a retrenchment. Resources are not being spent to develop procedures to comply with the mandated federal reporting requirements. "Evaluation of effectiveness" has been sidetracked and extensive new development of objective-referenced tests has come to a virtual halt in the evaluation of Title I in Michigan. It remains to be seen whether new state and local uses of the mandated evaluation models can be developed thereby reducing the conflict between federal constraints and state-local policy and program needs.

Summary

The role of the State Board of Education in Michigan changed dramatically in the last ten years. The transition was from a rather passive presence in Michigan education to an assertive leadership role. The leverage for the change was in large part because better data about the educational system became available to them.

The statewide educational assessment program was initiated in 1970. The data from the assessments were used as indicators of the weaknesses and strengths in basic skills education, and to monitor progress of the schools and districts of the state. The State Board based policy initiatives in compensatory education, accountability, equal education opportunities, and Department services to districts on information produced by the state assessments. The program became the "center piece" of elementary and secondary education in the state.

The assessment data were good indicators of needs, but were of very limited use in providing direction in dealing with the needs. A more indepth evaluation of programs was needed to identify "what works" to produce a better educational system and higher achievement for children, youth and adults. The State Superintendent, in recognition of this, centralized the evaluation function in the Department of Education, and over the years was most supportive of their work. The evaluations went through three phases; descriptive, success and effectiveness. Especially in compensatory education, the evaluation data were important in decisions of resource allocation, program management, and policy development. Whereas, the assessment data provided an indicator of problems, the evaluation data provided the data for addressing the problems.

Michigan education has come to appreciate the power of data in decision making. The State Board and State Superintendent appreciate the power of data and use it in forming policy. They have been better able to justify policy initiatives, and have been more assertive in taking initiatives to change and improve the educational system. Information from evaluations of the Michigan educational system has been used to form policies, and in turn, the policies have influenced the direction of the evaluation.

The proactive State Board and State Superintendent in Michigan used evaluation activities and data to establish a state presence in education during the 1970's. The tenor of the times was an "outcomes" orientation and a promotion of equity and equality for all children in education. The state accepted a responsibility for setting standards, for measuring impact, and for assisting schools toward improvement. This was a middle road between the policies in other states of setting statewide graduation standards based on competency tests, and leaving standard setting completely to local initiative.

FOOTNOTES

1. Michigan, *Constitution*, (1908), Article XI, Section 2.

2. *Ibid.*, Section 6.

3. Michigan, *Constitution*, (1963), Article VIII, Section 3.

4. J. Alan Thomas, *School Finance and Educational Opportunity in Michigan*, (Lansing, Michigan: Michigan Department of Education, 1968), p. 321.

5. *Ibid.*, p. 345.

6. C. Philip Kearney, "The Politics of Educational Assessment in Michigan," Paper Delivered at the Annual Meeting of the American Educational Research Association, Minneapolis, 1970, p. 5.

7. Thomas P. Wilbur, "Research Into the Correlates of School Performance: A Review and Summary of Literature," (Lansing, Michigan: Michigan Department of Education, 1970).

8. In Michigan, The University of Michigan historically did the accrediation of secondary schools on a voluntary basis. The Department of Education had neither the resources nor the inclination to take on this function.

9. Minutes of the State Board of Education, January 15, 1969, Department of Education, Lansing, Michigan, pp. 171-172.

10. Minutes of the State Board of Education, February 26, 1969, Department of Education, Lansing, Michigan, p. 223.

11. Minutes of the State Board of Education, April 23, 1969, Department of Education., Lansing, Michigan, p. 306.

12. Act 100 of the Public Acts of 1969.

13. John W. Porter, "Evaluating Educational Products." Speech at the Annual Meeting of the Chief State School Officers, November 13, 1972 at San Francisco, California, p. 3.

14. Michigan Senate, *Journal of the Senate*: No. 7, Regular Session of 1970, p. 83.

15. "Michigan Assessment K-12 District Correlation," (Lansing. Michigan: unpublished Michigan Department of Education staff paper, 1972).

16. *Ibid.*

17. *Ibid.*

18. Act 38 of the Public Acts of 1970.

19. Joint letter from John W. Porter, State Superintendent and Edwin Nowak, President of State Board of Education to Local Superintendents and Board Presidents, March 4, 1971.

20. *Local District Results, Michigan Educational Assessment Program: The Fourth Report of the 1970-71 Series*, Michigan Department of Education, Lansing, Michigan, December 1971, p. 171.

21. *Local District Results: The Fourth Report of the 1971-72 Michigan Educational Assessment Program*, Michigan Department of Education, September, 1972, p. 142.

22. Michigan Department of Education Memorandum, State Superintendent Associate Superintendents and Service Area Directors, August, 1974.

23. *Ibid*.

24. Michigan Department of Education Memorandum, State Superintendent to Administrative Council, April, 1977.

25. Stanley A. Rumbaugh, "A Review of the Status and Needs of Evaluation Activities in the Michigan Department of Education," (Lansing, Michigan: Unpublished Staff Paper, 1977), p. 1.

26. Michigan Department of Education Memorandum, State Superintendent to Associate Superintendents and Service Area Directors, August, 1974.

27. Michigan Department of Education Memorandum, State Superintendent to Administrative Council, April, 1977.

28. Stanley A. Rumbaugh and David L. Donovan, "Achievement Based Funding The Michigan Experience With Compensatory Education," Paper Delivered at the Nineteenth American Educational Finance Conference, Nashville, Tennessee, 1976, p. 5.

29. Rumbaugh, *op*. *cit*., p. 4.

30. Daniel E. Schooley, *et*. *al*., "Quality Control for Evaluation Systems Based on Objective-Referenced Tests," International Journal of Experimental Research in Education XIV, 2 (1977), 213-228.

31. Public Law 95-561, Section 183 (d).

32. *Ibid*.

33. Public Law 95-561, Section 183(f).

34. David L. Donovan, Daniel E. Schooley, "The National Title I Evaluation Model and Technical Assistance Efforts: A State Perspective," Paper Delivered at the American Education Research Association, New York City, April, 1977.

35. 45 CFR parts 116 and 116a, 44 FR 59152-59159, October 12, 1979.

THE INTERACTION BETWEEN POLICY AND EVALUATION IN STATE DEPARTMENTS OF EDUCATION

James H. Gold
GOLD Resources, Inc.
Madison, Wisconsin

INTRODUCTION

Although educational policy and evaluation have been a part of the American education system from its inception, the content, form and relationship between them have changed throughout the years. Early school policy was governed by concentration on the 3 R's and educational evaluation was based on the effectiveness of the individual teacher. In contrast, schools today have expanded programs far beyond the 3 R's in an effort to provide a more comprehensive education to greater numbers of students. The public has charged schools with the responsibility of addressing, and ameliorating, social problems. Accompanying this expansion of responsibility is an increase in public dollars from local, state and federal sources. As costs have risen and resources become less available, funding agencies such as private foundations, state education agencies, local education agencies, and the federal government have placed increased importance on promoting and funding educational activities that encourage desirable student behavior. Thus, program evaluation has become an increasingly important part of general school operations.

Specifically, the passage of the Elementary and Secondary Education Act of 1963 propelled the evaluation movement by requiring evaluation of State Educational Agency (SEA) and Local Educational Agency (LEA) programs supported by federal funds. This concept has grown horizontally to other federal programs and vertically to state and local programs. Subsequently, evaluation and policymaking have become an integral part of management concepts such as PPBS, management by objectives, and the more generic concept of accountability. Although variations exist in evaluation philosophy and application, a common theme is that feedback about behavioral changes resulting from program evaluations should be the basis for both policy and operation decision making at the program level.

The binding of program evaluation to policymaking is based on the notion that the scientific process which encompasses the evaluation process would solve the problems of educating American youth just as the same processes were able to put an American on the moon in the 70's. Thus, the expectations were raised that a one-to-one relationship could be developed between evaluation and policymaking.

Today, over a decade later, evaluators, educators, educational decision makers and other interested parties are faced with the reality that the one-to-one anticipated relationship has simply not evolved between policy and evaluation. Too often evaluation results are ignored in light of political expediency. Consequently, we are faced with the problem of improving utilization of evaluation information in policy development. This paper will detail some of the issues and present how one state, Wisconsin, is structured to relate evaluation to policymaking.

CONTEXT AND MODEL FOR EVALUATION AND POLICY

1. Contextual Factors

Prior to discussing the relationship between educational policy-making and evaluation, it is essential to understand some of the important contextual factors within which policy-making and evaluation operate in education. These factors are not new but are a reminder that the educational enterprise is dynamic, conducted by humans who possess the strengths and frailties which determine the outcome of all human endeavors. Systems are made up of people who should be accountable for results rather than for the failure of the "system" itself. Thus, the following contextual factors are presented as a framework within which most policy-making and evaluation take place.

Contextual Factor 1

Decisions in education are made by influencing those who have the final decision-making authority invested in them by state constitutions and laws.

Contextual Factor 2

Education is a political process which is strongly influenced by individuals and groups who are affected by the decisions. Their vested interest may conflict with the welfare of others.

Contextual Factor 3

Education is an enterprise in which people with diverse values must agree to live with a single set of policies and operations within an educational system.

Contextual Factor 4

Education is not an exacting science in which success or failure can be precisely predicted for any particular policy or program. Thus, a single best policy or program may fail to emerge.

Contextual Factor 5

Translation of programs and operations from policy may bear little resemblence to the intent of the original policy.

Contextual Factor 6

Both individuals and groups are often vying for limited resources, which often leads to conflict and competition rather than cooperation.

Contextual Factor 7

Evaluation conclusions are often contradicted and refuted by those desiring other outcomes. They may reinterpret data or present contradictory data which supports their viewpoint.

Because of these contextual factors we must begin with the assumption that educational decision-making regarding policy, programs and evaluation is not always rational, but is often based on influence and compromise. This political decision-making process is inefficient but successful in a multi-cultural democratic society in which public education is laid open, dissected, studied and gradually reconstituted in an evolutionary process.

Within this context we can begin to fit together the dynamics of policy and evaluation, realizing each state is somewhat different in terms of the power structure, values, and traditions which influence the policy/evaluation relationship. The importance of this paper is for the reader to gain insight into the dynamics of other states so that he may better utilize evaluation results in his policy development process.

2. A General Model

Educational policies are broad statements of intent which provide organizations with a basis for program design and implementation within a given State Education Agency (SEA). They provide a description of agency direction to those outside the organization. Characteristically, policies lack quantification, specific behavioral descriptions and program specifications. However, they usually do reflect a desired standard. For instance, a policy may be to "promote equal educational opportunities for all children in the state." This statement reflects a standard but does not indicate how it will be achieved or how one knows when it is achieved. These specifics are accomplished through the development of goals, objectives, programs and program evaluation.

The process of generating policies in an SEA is complex, and varies between SEA's. However, it does appear possible to develop a general model that reflects the dynamics of decision-making in most SEA's.

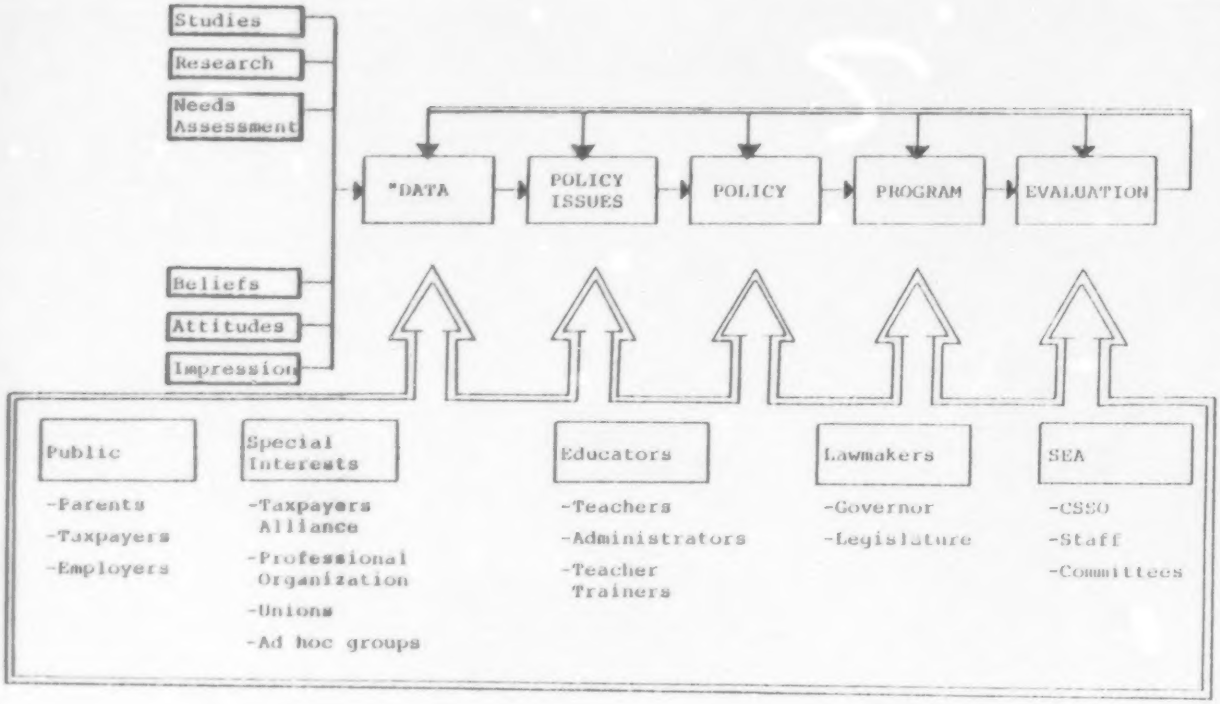
Figure 1 (page 6) is such a model and is designed to be flexible inorder to accomodate the variations which occur in SEA's. The general flow of the model shows that "data" raises policy issues, resulting in policy adoption, related programs and evaluation. The evaluation results are then used in revising policy, programs and even the evaluation itself.

The model begins with the disclosure of "data", which strongly suggests that either current policies be revised or new policies be developed. At the very minimum, the "data" raises serious questions as to certain unmet needs which must be addressed. "Data" in this model has two forms. First, people express their concerns based on their own experience as parents, educators, students, employers, and taxpayers. Although this data is neither systematically collected nor scientifically analysed it can have a powerful effect on educational policy, if a consensus of opinion is gained and the opinions are heard in a loud and clear fashion. This type of data can be more powerful in bringing about change than even the best evaluation studies.

The second type of "data" is the more systematic, consisting of test scores, surveys and research. Sometimes these studies are carried out to reinforce or change policy, while at other times their influence on policy is accidental.

Figure 1

GENERAL DECISION-MAKING MODEL



It is important to note that in most cases the two forms of "data" are used to complement each other. For instance, many people expressed concern over students not learning the basic skills. This notion was then reinforced by reports of declining test scores and other "hard" data. Thus, most state departments have developed a stronger and more visible policy concerning basic skills.

It is interesting to note that data has several points of entry into the policy development process. The chief State School Officer (CSSO), governor, legislature, state board of education and SEA staff are all viable candidates for influencing policy. Depending on the personal policies of each and their relationship to each other the entry points used are based on receptivety to change and the power to change. If policy change is desired it is most important to analyze the actors and select those who are receptive and willing to work for the desired change.

Once the data is revealed policy issues are determined by a number of different people. Professional groups, legislators, the governor, CSSO, and SEA staff offer various policy alternatives. Each of these groups may prepare issue papers, appear at public hearings or attempt to persuade others on an individual basis through rational analysis, emotional appeal or political compromise. The end result is either no policy, or a policy which is agreed upon by either the SEA, legislature, governor, or any combination of the three.

Policies are then translated into programs through the development of goals, objectives, activities and budgets. The legislature often goes beyond policy by determining some program specifications. Such activity can create conflict between SEA and legislators. It is at this point that lobby groups also work to insure that the funds are allocated and activities are designed to meet the needs of their constituency. Thus, these groups exert strong influence on both SEA's and legislatures during this phase of the process.

State programs are given varying degrees of autonomy in determining evaluation procedures. For some states program evaluations are required but the design is left up to the SEA. Other programs have the evaluation specifications spelled out either with general guidelines or specific activities. Regardless of the form, it appears that states are increasingly required to evaluate educational programs.

Evaluation results can influence virtually every phase of the model. Peoples' perceptions could be changed through the new information, thus the "data" base changes. The evaluation could cause a re-evaluation of policy or raise new policy issues. Certainly the management, organization, goals, objectives, activities and budget of a program could be affected. Whether any of these take place depends on the processes developed for handling data and the desires of those who control the data.

FEDERAL INFLUENCES ON EVALUATION POLICY

Since the passage of the Elementary and Secondary Education Act of 1963, the federal government has had greater influence on the policies and evaluations of state and local education agencies. The federal government of the 1960's began to intervene directly to overcome some of the large scale social problems of the decade. Education was among the many social programs affected. The major mechanism was the injection of massive funds to state and local education agencies for the purpose of designing and implementing programs that met the needs of society.

In handing out money for specific programs the federal government began to influence state and local district policies. For instance, acceptance of Title I funds increased state and local district's committment to improve the education of the disadvantaged. Although most states had some committment to this policy, the Title I funds increased that committment and made it very visible. Likewise, the original Title III greatly influenced policies for innovative programs and expanded the policy of publically documenting educational needs at a statewide level. Still other programs were responsible for implementing new management concepts and placeing an emphasis on educational planning on a statewide level. At the SEA level many, if not most, of the current central planning units had their origins established from ESEA funding. Thus, the federal government, through massive funding attached to specific programs, has had a great impact on educational policies at the state and local levels.

One of the key areas of influence is in evaluation of programs. A prime example is Title I, in which evaluations have gone from being locally designed to following more rigorous federally mandated requirements. At the outset, local districts were permitted great leeway in how their program evaluations were designed and implemented. The goals and objectives were strongly encouraged and the evaluation instruments were left up to the local district. As a result, evaluations ranged from excellent to totally inadequate. As the inconsistancy in quality became more evident, the federal guidelines for Title I evaluations were restricted to make them more consistant with good evaluation practice.

In addition, congress began to seriously question the effectiveness of Title I funds. This questioning led to changes in evaluation requirements and subsequently to evaluation practices. When faced with measuring the impact of Title I, educators were unable to aggregate Title I evaluation data from across the nation and gauge their effectiveness. Instead, case studies and anecdotal data were used to defend or attack the massive expenditure of funds. Subsequently, congress mandated that a method be developed to report on the impact of Title I to Congress.

As a consequence Title I developed four models for evaluation which generated data that could be aggregated at the state and federal levels. This strategy limited the evaluation instruments that were required (LEA's could supplement), the sequence of evaluation events, and, to a certain degree, the content required to be evaluated. In essence, the Title I requirements limit the required evaluation strategies, curricular content, and test instruments to those that Title I determine.

LOCAL INFLUENCE ON POLICY DEVELOPMENT

Local individuals and groups can influence policy decision-making through several mechanisms. Lacking definitive data on how effectively these mechanisms influence policy, the reader must draw upon his own experiences to judge the value of each.

1. Individual Personal Contact

In this case the individuals may call, write or meet with SEA management, legislators, or governor and express their opinions. For people viewed by government as opinion leaders, this type of contact is valuable. Otherwise, a large quantity of responses is needed to influence policy decisions.

2. Specific Issue Groups

There are ad hoc groups which pool energy and resources to change specific policies or procedures. Their interest is in a single topic and they use personal contacts, letter campaigns and media as their major modes of operation.

3. Task Forces and Advisory Groups

Historically, the governor, legislature and SEA have established task forces to review specific problems and make recommendations for policy and programs. These task forces are usually appointed by the governor and/or agency head, with the basis of selection not always clear. In most cases, the local constituency is represented. However, the individuals selected often hold views similar, or at least not incompatable with, the appointing authority. Thus many, but not all, task forces have a built-in bias.

4. Organizations

The local constituency is usually well represented by various professional organizations who have been increasingly involved in political lobbying efforts to serve the needs of their members. Teachers, administrators and business officials all have their representatives who monitor and influence educational policy and programs at the state level. In Wisconsin, some of these groups have formed an umbrella group with DPI for the purpose of discussing major policy and program considerations. Although the group has no formal authority, an overwhelming consensus on an issue, policy or program would have a great influence on those who make decisions.

In Wisconsin local influence has had a general effect of maintaining local control where the federal or state laws have not compromised it. In evaluation, most advisory groups opt for leaving the design and implementation up to LEA's and requiring as little extra work as possible to accomplish evaluation requirements. This position is partly due to the issue of control, but may very well reflect a feeling on the part of LEA's that evaluation data either is not, or cannot, be utilized enough to justify increased demands on the time, energy and money of the local district staff.

EFFECT OF STATE POLICIES ON EVALUATION

State educational policies effect evaluation in three ways. First, the policies may determine that no evaluation take place. This is usually accomplished by leaving the requirement for evaluation out of legislation and the budget. Thus, the programs are implemented and a general fiscal accounting is done, but no performance evaluation takes place.

Second, legislation and/or budget documentation may be very prescriptive in determining the evaluation policy and procedures. In such cases the evaluation requirements are often spelled out in detail regarding; 1) process, 2) instruments, 3) time-lines, and 4) reporting requirements. This situation places severe limitations on the SEA, but increases the probability that the legislature will have its evaluation policy carried out.

Third, the legislature/budget requires that the SEA evaluate specific programs that are being supported by state funds. The directive to evaluate is often vague, sometimes ambiguous and always open to interpretation as to legislative intent. Procedurally, the CSSO assigns the program to an individual to administer. The nature of the program and evaluation will then be determined by the personality, politics, professional persuasion, and program priorities of the program director and his/her superiors who have veto powers.

Like individuals, organizational units have their own personalities, politics and program priorities. Thus, central evaluation units tend to be interested more in performance based evaluations, using surveys and objective tests, while instructional people often emphasize process evaluation, using interview techniques or other methods which place less dependence on student performance. The nature of the evaluation, then, will be established by organizational assignments since an overall agency evaluation policy is often absent.

INFLUENCE OF EVALUATION ON POLICY

Educational evaluation has the potential for influencing four aspects of the educational enterprise. First is the establishment of programs based on evaluation of need. Once concerns are expressed as shown in the earlier decision model, the collection of systematic evaluation data may indicate the degree to which the concern is real. This use of evaluation could set the course for the content, process, and extensiveness of SEA programs. The results of these evaluations can directly effect the amount of fiscal and human resources made available to address the concerns.

Second, the management and operations of ongoing programs may be modified as a result of formative evaluation. These changes usually effect the activities of staff and students, but avoid any major changes in overall policy, missions, goals or objectivies. These changes are intended to promote more effective and efficient attainment of the original goals and objectivies.

A third aspect of evaluation is program monitoring, which is related to management and operations, but has the intent of insuring that the proposed program is the program being carried out. Program changes must be documented, verified and justified.

The forth area involves policy changes which determine the continuation of the program. Programs may be discarded because the evaluation shows them to be ineffective, inefficient or politically unnecessary. Evaluations may indicate that the goals and objectivies are both reasonable and based on desired standards. However, the end results simply may not meet the standards sufficiently to warrant continuation of the project. It is also possible that, even if the goals and objectives are being met, the cost in dollars and human commitment is too great for the outcome. Finally, when the evaluation results are placed in the larger context of total organization programming, other priorities may supercede the project as a result of changing needs and perceptions on the part of administrators and the public. Thus, the evaluation may contribute to the expansion, maintenance or termination of existing programs.

The final policy decision involves the choosing of one program approach over another, and the general significance of the evaluation results. As a result of a program evaluation an organization may choose to drop one program and adopt and/or expand a program that the evaluation has shown to be more effective, efficient, and/or politically acceptable.

In reality, evaluation studies have a relatively small impact on policies in comparison to their impact on program operations. This situation exists because of the nature of educational policy, the politics of the educational enterprise, and the current state of the science of evaluation.

As indicated earlier, educational policy is usually stated in broad and abstract terms which foster multiple interpretations of goals, objectives, programs and evaluation. Since most evaluations are designed to measure program objectives and activities, it is little wonder that policy is barely touched. In addition, most policies are robust enough to withstand significant program changes without requiring policy changes.

Statewide educational policy is usually very appealing to the public and appears to be in the same unassailable category as "chevrolet, apple pie, and the American flag." Within this context much public policy is traditional and insulated from rapid and extensive change. This stability is due to the balance of power between traditionalists, moderates and liberals who influence and make policy decisions.

This balance of power also explains, in part, the imposing role that the federal government and courts have played in bringing about both policy and program changes at the state and local level. During the last 25 years the government and courts have been liberal regarding social policy and programs, and have created policy and program changes at the state and local level by legal mandates and the infusion of large grants for educational programs. For example, statewide policies concerning equal educational opportunity, school desegregation, school finance, and library building programs were all changed dramatically because of federal intervention. Even curriculum policies and programs have been influenced dramatically by the creation of the National Defense Education Act (NDEA), Title III, and Title IV. It is important to note that the federal activity came about as a result of public concern and a national feeling that the policies and programs were necessary for the public good, not because a comprehensive evaluation concluded the changes were imperative. Society expressed concerns and the federal government and courts responded by establishing programs and policies that addressed them.

Evaluation has had a less than desirable effect on policy because the results are often inconclusive or contrary to previous studies. The technology of evaluation is not perfect and contains the bias of both the evaluator and the program staff which focuses on specific aspects of the program while minimizing and/or ignoring others. Even comprehensive evaluations have errors of both content and design which allow opponents to criticize the evaluation and discredit the results on techical grounds.

Similarly, education evaluation has not produced insightful, permanent, and significant discoveries which would revolutionize the enterprise as is true in other fields. Evaluation discoveries have no analogies to X rays, penicillin, or the electric light bulb. Instead, education has a series of fads such as MBO's, programmed instruction, and the open school, which loose their luster after a relatively short period of time or are refuted by contrary research within a decade. As a result, educators and evaluators have failed to create the public trust which other fields have developed through finding permanent and effective ways to address public concerns.

Finally, the traditional view of organizations having specific missions towards which all their human and fiscal resources are devoted is no longer appropriate. Factors such as limited resources, controversial policy issues, influence of special interest groups, public politics, organization politics and the openess of the democratic process have led to decisions that are less than optimal in regard to the organization's mission, but more practical in that they try to satisfy all the variables listed above so that programs and policies can be implemented. Thus, one should be little surprised when parents and others express dismay over their perception that the children seem to have been lost in the decision-making process while the survival of individuals and organizations appear to have been optimized. Although evaluators should be concerned about students, in reality evaluation results may be neither necessary nor effective when one or more of the above factors is given higher or even exclusive priority in decision-making.

STRUCTURE AND FUNCTIONS OF POLICY AND EVALUATION WITHIN THE WISCONSIN DEPARTMENT OF PUBLIC INSTRUCTION

1. Department Organization

The Wisconsin Department of Public Instruction is unique in that the state superintendent is both the major policy maker and administrator for education in the state. The CSSO is a constitutional officer elected in a popular non-partisan election every four years. There is no state board of education or any other state structure which supercedes the policy-making authority of the office. Thus, the CSSO is accountable only to the public every four years.

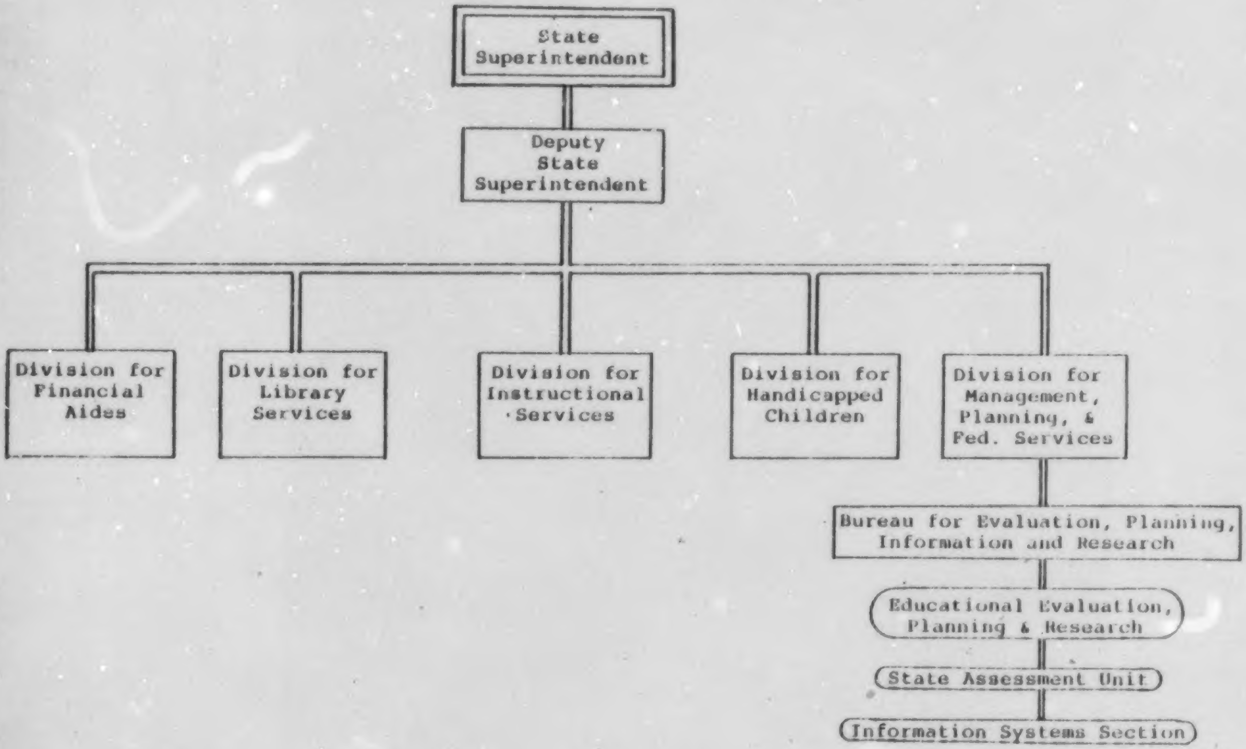
Implementation of policies and operations are accomplished in two ways. First, Wisconsin is under a biennial budget system with a budget review occuring on the off year. The budget is a mix of fiscal and program elements. However, the budget has been used increasingly by state agencies, legislature and governor for developing or changing policy.

The DPI creates a budget which is submitted to the Department of Administration for review after which the governor makes his recommendations to the legislature. Both budgets are reviewed and modified by committees and finally adopted as a total state budget. The governor does have line veto powers which can be overturned by a 2/3 vote of the legislature. This process generally follows the model described in the previous section and is greatly influenced by various special interest groups.

Organization of DPI, shown in figure 2 (page 19) consists of the state superintendent, an appointed deputy and five appointed assistant superintendents who serve at the pleasure of the CSSO. The department is divided into five major divisions including Financial Aides, Handicapped Children Instruction, Library, and Management, Planning and Federal Services.

Figure 2

DIVISIONAL STRUCTURE OF WISCONSIN STATE DEPARTMENT OF PUBLIC INSTRUCTION


State Superintendent
Deputy State Superintendent
Division for Financial Aides
Division for Library Services
Division for Instructional Services
Division for Handicapped Children
Division for Management, Planning, & Fed. Services
Bureau for Evaluation, Planning, Information and Research
Educational Evaluation, Planning & Research
State Assessment Unit
Information Systems Section

An Administrative Council made up of the CSSO, deputy CSSO, and five assistant superintendents review major policy changes. In addition, the CSSO confers with individual assistant superintendents, program staff, and numerous task forces and advisory groups for direction concerning policy and operations. However, final decisions regarding policy and implementation are the sole responsibility of the CSSO.

2. Evaluation Structure and Functions

Housed within the Division for Management, Information and Federal Services is the Bureau for Evaluation, Planning, Information and Research. This is considered the central evaluation unit of DPI and is divided into three units, two of which are involved directed in program evaluation. The unit entitled Educational Planning, Evaluation and Research consists of six persons who have varied responsibilities. Some design and implement evaluations for other DPI programs such as special education and nutritional education. Others review and monitor local evaluations for Title IVC projects. Some have been involved in a statewide needs assessment which had potential for statewide policy development. For the most part, the group is "on loan" to provide services to others who lack staff to fulfill their evaluation needs. This group also advises local districts on the design and implementation of local evaluation programs.

The second unit is the State Assessment Program which conducts an annual statewide assessment of pupil performance. This unit has a staff of six. In addition to statewide testing, this unit has been instrumental in developing a local option testing program and is in the beginning stages of developing an item bank to be used by LEA's.

3. Evaluation Methodologies

The basic methodology used in evaluation is that of establishing outcome objectives and measuring the attainment of those objectives for each program evaluation. In addition, some process evaluation may take place to insure that the program proposed was in fact the program evaluated. Achievement tests, interviews and questionnaires have all been employed as data-gathering tools.

4. Problems and Constraints

Lack of clarity concerning what clients want from evaluation

Many clients come to evaluation without much knowledge of what questions they want the evaluation to address. Subsequently, a great deal of time must be spent on clarifying goals and objectives. To some clients this is both a tedious and often threatening task.

Attitude

Some clients are afraid of evaluators because they are intimidated by them and/or feel they are being personally evaluated. Such feelings hinder the evaluation process in that people either become resistant to the process or agree to things they later reject under the pretense that they did not understand them.

Minimal Effort

Clients are often being forced into evaluation to hold on to their funding source, and, therefore, desire to do the minimum to meet the requirements. They are just going through the motions.

Communications

Evaluator and clients often fail to communicate ideas and assumptions, thus a common understanding does not exist. Such communication gaps are caused by both language differences and varying degrees of receptivity to ideas and viewpoints on the part of both the evaluator and client.

Use of Evaluation Results

After the time and money are invested, after fulfilling an obligation to evaluate, clients do not utilize results adequately. There is resistance to pre-planning the use of results and a propensity for a "wait and see" attitude. However, most evaluators do believe that results are used in a less systematic and less visible fashion.

Difference in Philosophy

Differences in evaluation philosophy exist among educators, especially in terms of the "hardness" or "softness" of data required for adequate evaluation. In the Wisconsin DPI, as in most agencies, there are those who place greater emphasis on process rather than outcomes. Thus, conflict arises as to how much reliance should be placed on student performance data vs. other types of information. If an evaluator who is a strong believer in performance data is "loaned" to a program whose personnel tends to believe more in process or other data, the evaluator is constrained in applying his profession as he/she believes is appropriate.

In addition, the CSSO directed that public involvement be significantly increased in all phases of the assessment. This change resulted in the assessment of practical skills and knowledge in addition to the purely academic objectives the assessment had previously focused on.

The changes in policy and procedure were consistant with the general policy of the CSSO regarding local control. The assessment was to be developed for, and by, the public with technical assistance provided to local districts on a volunteer basis. The assessment design was a "safe" compromise which showed DPI was concerned about accountability and yet was not demanding enough to pose a direct threat to any particular special interest group. In summary, the assessment was a reactive program designed to quell concerns over student performance. Its purpose was to provide public information and not to comprehensively evaluate specific educational programs in Wisconsin.

Although the yearly results made interesting newspaper coverage, it is difficult to identify if the assessment has been instrumental in changing policy. No new programs have been initiated and no new funds have been generated by the assessment results. There is scant evidence that the education of Wisconsin children has improved, or that any improvement, which may have occurred, would be due to the assessment. On the positive side, there is evidence that local districts using the local option program, have used it for program improvement. This effort is encouraging, since districts are attempting to find ways to make evaluation figure more prominently in policymaking.

PUPIL ASSESSMENT: A CASE HISTORY

In 1971 the Wisconsin legislature with the support of the DPI enacted S.115.28(10), which mandated the department establish a pupil assessment program within very broad guidelines:

> Develop an educational program to measure objectively the adequacy and efficiency of educational programs offered by public schools in this state . . . Assessment shall be undertaken at several grade levels on a uniform statewide basis. (S.115.28(10)

Unlike other states, this legislation passed with a minimum of debate and little organized opposition. It received neither wide press coverage nor special interest group attention. Most importantly, the legislature did not provide any state funds for initiating the program, which may account for the lack of interest among law-makers and educators regarding the passage of the bill. Thus, the DPI was mandated to provide a program of pupil assessment, but was given little legislative guidance, interest or funding.

Fortunately, the state superintendent was committed to the concept of accountability and therefore allocated discretionary federal funds for starting the program with the intent that the state would eventually take over its support. This became a reality in the following biennium, when the state took full support of the program. The state increased allocation for the program each biennium since.

From 1971-75 the assessment program developed a set of goals for education, and implemented assessments in reading, math, science and social studies. The Eleven Goals for Education were intended to be the basis of the assessment and were to provide direction for education in Wisconsin Public Schools. The test instruments were all objectively referenced, and put together by Wisconsin educators, or selected from National Assessment of Education Progress (NAEP). They were administered to a random sample of public school students based on a two-stage random sample design.

During the first four years, several events did and did not take place. No definitive purpose for the assessment was established or documented. There was little consideration given to the nature and type of data to be collected. There was little consideration of persons responsible for content and technical quality of the instruments. No one was designated responsible for the content of the reports. Thus, an internal tug-of-war began between assessment personnel, the specialist, and management of other divisions. As a result, until 1976, products were deemed unadequate by the Assessment Director. Many policy-makers, outside of the DPI, questioned the value of assessment data.

At the end of the 1975 Assessment, an internal evaluation of the assessment program was done by the assessment staff and a Technical Advisory Committee. The evaluation concluded; the sampling procedures were excellent, the logistical systems for administering the program were excellent, the test instruments needed some refinement, and the program did not seem to meet the need for student information. Thus, an assessment program had been created that operated well, but did not satisfy needs of educators, law-makers, or the general citizenry of Wisconsin.

As a consequence, the CSSO directed the assessment staff to accomplished the following:

1. Provide the citizenry of Wisconsin with a statewide profile of the quality of education as reflected in students' ability to demonstrate expected knowledge, skills, and attitudes.
2. Provide state officials with student performance information for use in educational policy development and/or communicating with constituents.
3. Provide state officials and the general citizenry with a profile of Wisconsin pupils' performance as compared to a national average.
4. Provide school districts with the opportunity for self-evaluation, using the methodology and products of the Wisconsin Pupil Assessment Program.

HEURISTICS FOR INTEGRATING POLICY AND EVALUATION

For the prupose of clarity it is necessary to draw a distinction between local project evaluations with limited local implications, and statewide evaluations which have potential implications for the entire state education system. Policy questions are more easily addressed in projects. The appropriateness of a decision can be judged against the concerns and needs of a relatively small group of people instead of judging whether a paticular policy or program is appropriate for 436 LEA's who have both common and unique needs. Consequently, policy may be more controversial at the statewide level than in a single LEA.

The sheer size of a statewide evaluation and decision-making process opens complicated political channels which are often difficult to control. The power structure may shift with the policy issue in statewide processes, whereas, at the local level the power structure appears to be more stable.

Statewide evaluations are usually carried out by SEA staff, while local project evaluations may involve SEA staff, local staff or outside consultants. Since statewide evaluations require an internal evaluator, his/her role may be different than an outsider's. The internal evaluator begins with a particular status in the organizationand is, in all likelihood, less prone to deviate from the organizational norms of communication, attitudes, and innovation. This individual is already part of the organizational structure and, in all probability, has been pigeon-holed into specific political and philosophical categories, making it less likely that the organizational staff will view the evaluator as unbiased. The SEA evaluator must then make a conscious decision to either facilitate the evaluation process by providing technical expertise, or take an active role in the politics of the situation, thereby influencing the design, implementation and utilization of the evaluation.

1. Project Evaluation

A. Make sure your credibility with the organization is established by providing documentation of work history and references from other evaluation projects you have conducted. Submission of an evaluation report which you have completed would also be valuable.

B. Determine at an early state both the formal and informal structure for decision-making in the organization, and deal with the appropriate decision-makers regarding the evaluation design and implementation.

C. All communications should be done at a language and technical level appropriate for the audience. Do not use the average level but the lowest level in the audience so that you will communicate effectively with all people involved in the process.

D. Be prepared to deal with people on a simplistic level. Assume responsibility for increasing the audience's knowledge of evaluation techniques and of how results can be utilized.

E. Clarify the roles and responsibilities of the evaluator, staff and administration.

F. Do not go into an evaluation situation with preconceived notions of what the evaluation should do or how the evaluation will be conducted. Be sensitive to local needs vs. textbook models. In the final analysis, the best evaluation design will be useless unless it meets local needs.

G. Do not take sides on local political issues, but act as a concillator in bring about compromise. Make suggestions and relate relevant research, but do not impose your viewpoint.

H. Identify the purpose of the evaluation and the specific questions that need to be answered. This is essential since most clients do not understand what an evaluation can do or what they want. Be honest about what questions can and cannot be adequately addressed, and explain why. State how much staff time and what resources are required for answering questions. Do not let the client expect more than the evaluation or evaluator can deliver. Document the purpose of the evaluation and the specific questions to be addressed and have both the client and evaluator sign it.

I. Develop the evaluation process plan. Include activities, responsibilities and timelines. This may be accomplished through developing options which the client decides from, based on your analysis. This decision should be a team process so that the client will feel ownership and try to tailor the evaluation to the local situation.

J. Develop an analysis plan that specifies how the data will be translated into answers to the evaluation questions. Avoid statistical jargon. Present the plan so that everyone involved will understand the process and the final outcome.

K. Develop an interpretation and utilization plan that delineates who, how, and when the results will be interpreted. This step is critical in utilization, since most clients are inclined to let the results determine the utilization of data. By defining how specific desirable and undesirable outcomes will effect policy and operations, the client will probably be more committed to following through with specific actions.

L. Develop a dissemination plan which targets evaluation results and recommendations to audiences in a form that they will read and follow through on. Produce technical and summary reports which convey the same information in different forms. In some cases dissemination may utilize alternative media such as transparencies, television or slide/tape presentations.

M. Keep key decision-makers informed as to the progress of the evaluations and any unusual findings. Do not drop surprises on the client to which he/she is not ready to respond.

N. Make sure the role and responsibility of the evaluator is clear in terms of information release. Do not release information without the clients approval. Requests for information should be directed to the client.

2. Statewide Evaluation and Policy Interaction

Most statewide evaluations are conducted by the SEA staff, thus, there is little client/evaluator conflict. However, the evaluator must interact with other government and public entities, finding the most resistance in intra-agency dealings. Thus, the heuristics presented below assume evaluation and policy interaction are internal agency activities that may involve extra-agency politics.

A. It is essential that the evaluator have the full confidence of the agency's top management. Since statewide decisions are made by the CSSO and/or state board, it is imperative that the evaluator build a strong track record with them. He/she must produce evaluations that fit their expectations and are politically astute, flexible and technically competent.

B. The evaluator must build a positive relationship with those staff members who may be affected by evaluation outcomes. Agencies are generally resistant to change. Individuals may be directly threatened by evaluation, and may do everything in their power to overtly and covertly impede the evaluation effort.

C. Evaluators should build relations with influencial groups outside the agency and educate them as to how evaluation results can be used in making decisions. These influencial people include budget analysts, legislative staff, professional organizational staff and media people. Evaluators can educate such people by holding conferences and workshops for them, or by meeting with them individually to inform them of evaluation progress. Under no circumstances should such meetings take place if they conflict with an agency policy or rule. Likewise, the evaluator should not promote ideas which are contrary to agency policies, rules or regulations. The intent is to educate and build confidence in the evaluation process.

D. The purpose and objectives of any statewide evaluation should be clearly delineated and approved by the CSSO and the evaluator. The evaluator should offer alternatives and recommendations as to what questions the evaluation may address, as well as, an analysis of the policy and operation implication of each.

E. An evaluation plan should be developed that includes activities, responsibilities, and timelines. This plan is then signed-off by the CSSO and management staff who are involved. The signing of the plan represents a commitment to implement it. Management must insure that staff members carry out the plan even if it means reduction in other staff activities.

F. An analysis plan should be developed to specify how the data will be translated into a format that will answer the questions addressed in the evaluation. This plan provides a mechanism for communicating what will and will not be done with the data.

G. A utilization plan should be developed to indicate how the evaluation results will be used. Positive and negative results should be analyzed and potential actions described. Such a plan represents a public commitment to use data in specific ways. It may prevent the collection of extraneous information that is costly and inconvenient.

H. Provide a dissemination plan that directs evaluation results and recommendations to target audiences in a form that they will read and follow through on. Produce technical and summary reports which convey the same information in different forms. In some cases, dissemination may utilize alternative media such as transparencies, television or slide/tape presentations.

I. Have evaluation results reviewed by professional and lay people who have different points of view. Provide mechanisms for each interaction with the two objectives of, 1) having people with diverse viewpoints gain a better understanding of each other and the issues, and, 2) obtaining new perspectives on the issues. Such an open dialogue promotes common understanding and increased support for subsequest actions.

Although the heuristics described above may aid utilizing evaluation results in policy-making, the dey to success is in the personalities and politics of individual decision-makers. It is obvious that some decision-makers have the confidence and ability to attempt revolutionary changes while others are satisfied to let the "system" evolve at its own pace. Regardless of any particular situation, the evaluator must understand that evaluation results will always be used in a political context. Unless attitutes change, most systems will decide to provide sufficient, not optimal resources to an organization in helping it attain its goals. Credible evaluation, then, is essential for verifying needs and effectiveness of programs within the broader context of decision-making.

EVALUATION AND DECISION MAKING IN WASHINGTON STATE

Alfred Rasp, Jr.
Washington State Department of Education

BACKGROUND

On January 11, 1973, a program evaluation section was officially established in the office of the Washington State Superintendent of Public Instruction; and for the first time, at least in modern history, an emphasis was placed on the measurement of program impact. This does not mean that previous superintendents lacked interest in the success of programs, but it did express a new concern for generating evaluative data as a basis for policy making. This paper will attempt to describe both the organization changes that have taken place and the interface between evaluation and decision making in Washington state.

To make sure there are no misunderstandings about intent, names or geography, three ground rules will be established. First, this description of events will be neither an expose' of agency practices nor a positive self-serving statement lauding the efforts of the evaluation section. In the words of Howard Cosell, the goal is to "tell it like it is." Second, in addition to the use of the standard educational acronymes such as LEA, SEA, and USOE, the Washington State Superintendent of Public Instruction will simply be abbreviated to SPI in the name of economy. Third, to avoid misunderstanding whenever the word

Washington appears singularly, it will mean "State of" not "D.C." People in Washington state just prefer it that way.

Dr. Frank B. Brouillet was elected SPI in the fall of 1972 and officially launched his administration January 11, 1973. His professional career represents an interesting blend of education and politics. He has professional experience from both the school and college levels. He is a former teacher, counselor, coach and administrator. He has degrees in economics and education and an earned doctorate. Perhaps most unique in this background blend--he served 16 consecutive years in the Washington House of Representatives and consistently provided legislative leadership in educational affairs.

This combination of experiences has led Superintendent Brouillet to a three-part educational philosophy. He professes a firm belief in the importance a local control, a commitment to providing the resources necessary for a quality education and a dedication to the basic tenents of educational accountability.

It is the third element that is of special significance to this discussion of evaluation and decision making. Being an insider to the working of the legislature, Brouillet knew long before being sworn in as SPI that perhaps the only way to expand the amount of state resources for education, and at the same time protect and strengthen local control, required close attention to accountability. He knew that maintaining or increasing the financial support for programs in existence and initiating new programs depended in large part on providing the legislature assurance of the following: first, that the program is necessary--that the need really exists; and seccnd, that the impact of the program can be measured, and, of course, that the results are positive. The methodology of evaluation plays a central role. The key question, however, is not whether one alternative or treatment is more efficient or effective than another, but in a more basic sense, does the alternative selected make a difference? Is there an impact? Assuring need and effectiveness become prime concerns for evaluation in the political accountability system. The influence of an elected superintendent with an educational and legislative perspective clearly makes an impact on evaluation practices.

ESTABLISHING AN EVALUATION SECTION

The rhetoric of the campaign trail, increasing programs, and protecting local control by establishing need and measuring impact became criteria for establishing a new section within the office of the Superintendent of Public Instruction. This program evaluation section gave concrete, visual proof that the new superintendent meant what he said--there was clearly a place to point to on the organizational chart. (An organizational chart appears at the end of this chapter.)

It should be noted that in many respects the development of the evaluation section was a process of putting "new wine in old bottles." Because at the same time the superintendent took office, the legislature in the name of efficiency placed a limit on the staff size of state agencies, and no trained evaluators were hired. The evaluation section was formed with personnel on hand, only the responsibilities were new.

As time passed, the deputy superintendent was instrumental in shaping the section into an effective work unit and expanding the emphasis on program evaluation. He recognized early that in the name of objectivity program managers should not evaluate their own programs and that outside contractors could not interact favorably with the legislature. With his leadership, several programs were designated as priorities and small amounts of their administrative funds used to establish project employment positions in the evaluation section and to hire staff to carry on the evaluation activities. This move gained both objectivity and credibility, as well as the required evaluation data.

The title of the section has changed during the years to reflect new emphases and to better reassure the legislature and the public that the accountability charge is being carried out. In the beginning, the name Program Evaluation seemed to be the

124

answer. It soon became apparent that not having a "research" descriptor appear in the organizational roster was causing the agency to miss important contacts. Thus the section title was expanded to Program Evaluation and Research.

By the mid-seventies, however, evaluation in the SEA setting had generally subsumed research activities and, with the advent of "golden fleece" awards and other indicators of the public's low esteem for educational research, the title was changed to Testing and Evaluation. This choice dropped "research" and added emphasis to the more popular notion of "testing." At a time when legislative debate on questions of testing was long and loud, the reference to testing in the section title reflected the SPI's intent to meet issues head-on.

A broadening of the accountability concept took place in the mid-seventies when the Seattle School District successfully sued the state for not meeting the constitutionally mandated duties "to make ample provision for the education of all children...and provide for a general and uniform system of public schools." This legal battle led directly to the passing of the Basic Education Act (BEA) in 1977 and to the need for "evaluating" LEA compliance with the provisions of the law. At the same time, the State Board of Education renewed its interest in expanding the concept of school accreditation to focus on faculty self-evaluation, that is, on improvement through evaluation. How better could the SPI meet these accountability challenges than by adding the responsibilities to the Testing and Evaluation section and changing the title to Testing, Evaluation and Accountability? In 1980 the routine compliance checking activity of the Basic Education Act will move out of the evaluation sphere to a more appropriate long-range setting. Perhaps the section title will stabilize as Testing and Evaluation.

This discussion of names may sound superficial, but it is important to note that the activities of the section have always received the necessary financial support. Hopefully, the major reason for this fortunate circumstance is that the section staff has discharged its assigned responsibilities with professional competence. The changing of titles, however, does reflect an attempt to match the "mood of the times," certainly that of the legislature, and to build confidence in the SPI's intent to establish program need and to measure impact as a decision making base.

CURRENT RESPONSIBILITIES IN TESTING, EVALUATION AND ACCOUNTABILITY

Testing

Currently, the major testing responsibility is to carry out the mandates of the state testing law passed in 1976. This law, titled Student Achievement Surveys and Tests, requires state testing and reporting at three grade levels. The SPI must annually administer a standardized achievement test in the basic skills of reading, language arts and mathematics to all fourth grade students. The results [illegible] the testing, along with the relationship of achievement to appropriate input variables, are to be reported to the legislature, LEAs, and subsequently to the parents of children tested so that parents can compare the achievement levels of their children with others in the district, state and nation. In grades eight and eleven, samples of students sufficiently large for generalizing to the entire state, approximately 2,000 at each grade level, must be tested in reading, language arts and mathematics and the results reported to the legislature at least once every four years. The law also encourages local school districts to conduct diagnostic testing in grade two but does not assign that responsibilty to the SPI.

The main intentions of the legislature in passing the law are clear. There was first an interest in ascertaining the impact of basic skills instruction. This was typified by the questions: What are the achievement levels of Washington students? How does Washington performance compare with the national? Do areas of weakness requiring special attention exist? A second purpose led to the display of district summaries of fourth grade test results. The interest being two-fold--to spotlight high achieving districts in order to learn from their success and to isolate low achieving districts for special assistance. A third purpose was to provide parents and the public information about the impact of schooling, that is, to encourage educators to more fully share information related to program outcomes.

The law is implemented through heavy reliance on contracted services. To accomplish major tasks such as the printing and scoring of tests, logistical services and analysis, requests for proposals are prepared and sent to interested bidders. The technical proposals submitted are reviewed by outside panels of experts working independently. The recommendations of the technical review panels are supplemented by the SPI staff analysis of bid amounts; the superintendent makes the final decisions, and contracts are written with successful bidders. In Washington, contracts for $2,500 or more require that a competitive bidding process be used. Single source contracts for larger amounts must be justified and defended. In the case of contracting with other state agencies, for example, universities, educational service districts and LEAs, waiving the competitive bidding process is not difficult; however, when agencies other than those of the state are involved, great care is taken to explicitly follow the rules.

Since the total professional staff responsible for the testing activity is less than one full-time equivalent, contracted services are necessary and play a crucial role. The typical pattern is one in which large contracts for specialized services are awarded on the basis of technical merit and competitive bid. The assistance of additional personnel is gained through contracts with the other state agencies or school districts. Specific tasks are completed occasionally through the use of single source personnel service contracts under the $2,500 amount. Developing work plans and time schedules, preparing requests for proposals, reviewing bids, writing and managing contracts are necessary skills for administering the Washington testing program.

The testing results are reported in several forms and through several channels. In the case of grade four individuals, classroom, school, and district level reports, including summary data and item analyses are delivered to the LEAs as soon as possible after the October testing. In December, the state's performance is publicly released to the LEAs and media. By the end of February the State General Report and District Level Summaries is disseminated to the legislature, LEAs and media. With the sample studies at grades eight and eleven, there is less information to report. Since a sample is used, no classroom, school or district reports exist. When possible, individual student results are returned to the schools, but the reporting consists primarily of a news release of the state's results followed by a general report sent to the legislature, district superintendents, principals of schools with the grades tested and the media.

An additional thrust of the testing program aims at helping personnel in local districts to improve their skills in selecting, administering, interpreting and reporting test results. This effort usually takes the form of workshops conducted throughout the state. The first series, timed before the October testing of fourth grade students, focuses primarily on test administration. A second series, conducted after the state's fourth grade test results have been returned to the districts, emphasizes interpretation, reporting, and use of test results for instructional improvement.

Evaluation

Major evaluation efforts revolve around the evaluation of selected, priority program. These are the programs in which the SPI has a special interest because they involve large sums of money and/or are compensatory or categorical in nature and/or are politically sensitive.

For the past two years, evaluation priority has been placed on six programs: Title I, Title I-Migrant, federal programs for the handicapped, the Washington Urban, Rural, Racial Disadvantaged program, educational clinics for dropouts, and the Title IV, Part B learning resources program. Although these are all designated as important, the evaluation responsibilities vary from program to program, with greatest efforts in Title I and Migrant, and least in the areas of Title IV, Part B.

In both Title I and Migrant, full annual evaluation reports are prepared for USOE. These reports are based on the computerized aggregation of data from applications, monitoring forms, interim reports, and year-end reports, as well as fiscal files and program office files. The annual evaluation reports describe how the program resources were used, what outcomes resulted, what trends developed, and what special problem areas existed. The annual reports also show the extent to which the state plan goals and objectives were met.

In addition to the preparation of the evaluation report, the computer data files are summarized and printed to provide periodical management information for the program staffs. During the course of the year, the evaluators also assist in training LEA personnel to use program forms and procedures.

There are two points of emphasis related to the evaluation of federally sponsored activities for handicapped students. For several years, the main responsibility was for evaluating the special state projects provided by federal discretionary funds. This was accomplished through year-end report data and on-site reviews. More recently, with the impact of P.L. 94-142 and the mandated individualized educational programs for handicapped students, the evaluation staff has been working primarily to assist in the development of a computer processing system for management information, including an emphasis on organizing monitoring and evaluating data.

The Washington Urban, Rural, Racial, Disadvantaged program was expanded in 1979 to include Remediation Assistance (RAP). Whereas the regular URRD program has provided money for a wide range of crisis oriented projects for the past decade, the RAP addition is strictly a compensatory program modeled sufficiently after Title I to qualify Washington for the Title I incentive grants when the federal funding becomes available. The evaluation section involvement with URRD takes many forms, including: the review of the evaluation plans specified in the grant applications, onsite project evaluations, the computer aggregation of compliance monitoring data, the follow-up study of students served, and a computer summary of application data. In the case of the new RAP component, assistance has been given in the development of program guidelines for LEAs and in the preparation of the reporting documents. The year-end evaluation activity will include the preparation of a statement on achievement gain in the style of Title I.

Evaluation assistance to the manager of the Title IV, Part B learning resources program typically has taken three forms. The application and financial data are stored, aggregated and tabulated by computer for both interim management information and year-end reporting purposes. The results of LEA compliance monitoring by learning resources program staff are entered into the computer and aggregated. Three to five case studies involving onsite reviews of LEA activity were prepared by the evaluation staff during each of the past three years. At the end of the year, computer printouts of updated program information and monitoring reports, along with draft copies of the case studies, are delivered to the program manager, who is responsible for preparing the annual report for USOE.

For the past two years, the State Board of Education has been required by legislation to certify education clinics organized to provide programs for school dropouts, and the SPI has been required manage the funding process and evaluate the programs. Because of the special legislative interest, the activities are politically sensitive beyond the small amount of money involved. The law itself calls for the evaluation of superior performance based on educational gain as related to the difficulty of educating the students and efficiency in terms of per pupil expenditures. The demands for evaluative precision outstrip the current state of the art. An achievement and superior performance report is prepared annually based on data aggregated from individual student record forms that are submitted by the clinics for each student entering and exiting the program. From this information, a description of each clinic is prepared showing a difficulty to educate factor, an achievement factor, and an efficiency factor.

Accountability and Other Responsibilities

Since the Basic Education Act went into effect in September of 1978, school districts must be judged in compliance, or have certain regulations waived, by the State Board of Education before the SPI can distribute the funds provided by the legislature to them. With 100 percent of the funds for the basic program moving to districts through these channels, the determination of district compliance is of crucial importance. For the past two years, this responsibility has been fulfilled by the Testing, Evaluation and Accountability section. Forms are developed and distributed; reports reviewed. Recommendations based on the district input are presented to the State Board of Education for action. As the board judges districts to be in compliance, the SPI's Division of Financial Services manages the apportionment of funds. Approximately one billion dollars flow to Washington's 300 school districts through this process each year.

The school accreditation programs are also administered by the Testing, Evaluation and Accountability section. In Washington, two accrediting programs interface. The State Board of Education by law must provide an accreditation process to any school that applies. The program is voluntary, not a basis for funding available to all schools. Although the State Board has accredited secondary schools for years, in 1979 the law was amended to add elementary schools. The accreditation program is currently in the developmental, field test stage, and the goal is to have a process operational by September, 1980. The second accreditation program is that of the Northwest Association of

Schools and Colleges. This program is managed by the section director who serves as the chairman for the Washington Committee of the Commission on Schools and as a Trustee of the Association. The Northwest secondary school accreditation process has operated in Washington since 1917 and currently involves approximately 180 member schools. Management has been a responsibility of the director since he joined the SPI staff in 1970 and a responsibility of testing and evaluation since the section was formed. Educational improvement is the goal of both programs. The central elements of each revolve around determining that the resources required for a quality educational program are present and conducting an indepth self-study with external verification. The state board activity as part of the SEA activities may move to another section in 1980.

Section personnel are responsible for a number of other activities--some of which are closely related to testing and evaluation and others only by a great stretch of the imagination. For example, liaison is provided with a number of organizations, including: American Educational Research Association, Washington Educational Research Association, Northwest Evaluation Association, National Assessment of Educational Progress, Northwest Directors of Assessment, Committee on Evaluation and Information Systems, Region X-Title I Technical Assistance Center, Northwest Association of Schools and Colleges, National Study of School Evaluation, Association of Washington School Principals, Elementary School Principals Association of Washington, Washington Junior High/Middle School Principals Association, Washington Association of Secondary School Principals, High School-College Relations Committee, Washington Pre-College Testing Program, and the Washington Alternative Learning Association.

Questions dealing with correspondence schools and high school graduation requirements are also answered by section personnel. In addition, the section is the clearinghouse for many research activities in areas related to testing and evaluation. A current example is the "High School and Beyond" study being conducted by the National Opinion Research Center for the National Center for Educational Statistics. A project just concluded provides another example. From July 1977, to the fall of 1979, the Northwest Reading Consortium, a four-state Research and Development Utilization Program established by the National Institute of Education, was coordinated by the section. Section staff also provide technical assistance in planning, testing, evaluation and research as requested within the state agency and outside.

STAFFING

During fiscal year 1980, several staffing patterns will be used to provide the human resources necessary to complete the assigned responsibilities. At this time there are eleven people regularly on the section's payroll. Seven of these are professional educators and four are secretaries.

The seven professional staff members have all been teachers, but their backgrounds vary greatly. Five have earned doctorates, and they bring great diversity to the section because each studied a different specialty. For example, one received the degree in curriculum and instruction, another in educational psychology , a third in counseling and guidance, a fourth in reading, and the last in the administration of higher education. Of the seven, two are former principals, one a school counselor, one a school psychologist, one a reading specialist, and one a former state education association president.

As a result of experience and graduate study, all have backgrounds in educational research, but none have extensive formal training in evaluation. This is not to say that there is a lack of expertness. Since the section was launched in 1973, steps have been taken to develop the required skills. Through individual initiative, section staff development activities and on-the-job training, the staff has gained a high professional level of competence. In the areas of large scale assessment, program evaluation and the use of the computer to facilitate the aggregation of evaluation data, the professional strength of the section is noteworthy. It should also be noted that the section has earned a good reputation--a high level of credibility even though it frequently deals with tough topics that are not always viewed favorably by LEAs or by others in the SEA.

The seven professional staff members represent two hiring patterns and three funding sources. Six of the seven are regular civil servants. The seventh is hired on the basis of special project need, and the employment must be renewed and approved at the beginning of each project year, depending on the availability of funding. The section budget is based on three sources of funds: state money provided by the legislature for the testing progam and for general SPI activities, such as administering the Basic Education Act and State Board of Education's accreditation program, federal dollars for state leadership in education and small amounts from the administrative money of Title I, federal handicapped, state compensatory and federal learning resources.

The regular secretarial staff consists of four people. There are two full-time secretaries, one part-time secretary working on program evaluation reports, and one part-time secretary assisting in the computer processing of data. As overload situations arise, temporary help is added as required with a minimum of bureaucratic strain.

Eleven people cannot attain all of the objectives flowing from the many assigned responsibilities, but with the size limitations imposed by the legislature, there are not staff years available for hiring additional permanent personnel. Steps are taken to augment the staff through the use of personal service contracts. In some instances, the contracts call for another agency to provide personnel who will work under the direction of the section. In other cases, the accomplishment of specific tasks forms the prime objective of the contract. Occasionally when the need for assistance is short in duration or the specific task is small in scope, a personal service contract may be negotiated directly with an individual.

Using contract has both advantages and disadvantages. Certainly control over the size of the permanent staff is maintained, and there is an efficient flexibility for peak load staffing. However, the negotiating, writing, defending, and managing of contracts is time-consuming and frequently calls for efforts over and above the time normally spent on the superivision of personnel. There is also a potential problem in the lack of staff continuity and commitment to long-range goals. Because of the different types of contractual arrangements used, it is difficult to estimate the number of full-time equivalent staff members that serve the section during any given year.

RELATIONSHIPS WITHIN THE AGENCY

There are a number of relationships with the SEA that help to define the roles, the responsibilities, and, in a sense, the location of the evaluatin unit.

Success in fulfilling the evaluation responsibilities depends on close and positive working relationships with SPI program managers. The evaluation staff and program staff negotiate a work plan specifying the activities, timelines and staff responsibilities that will guide the evaluation effort throughout the year. The cycle of involvement typically begins with a review of a program's state plan. Placing emphasis in two areas, the objectives are analyzed to ensure that they reflect the major intended outcomes of the program, and the evaluation plan itself is elaborated and brought up-to-date. Application and reporting forms are examined to make sure that they will provide the information required and provide it in a condition compatible with computer data processing. With the assistance of the computer, information is aggregated and reported to managers on the predetermined schedule. The outline for the final report is discussed with the program staff, and draft copies reviewed before final printing. This review is conducted to provide program staff an opportunity to point out possible data errors and to provide a first-hand knowledge of the contents before the report is disseminated.

In order to promote objectivity, the evaluation and program activities are clearly separated by housing each in different divisions of the LEA. None of the programs for which an annual evaluation report is prepared are located in the Division of Instructional and Professional Services--the home of the evaluation unit. This separation solves two problems. Since the program staff is not evaluating itself, there is an appearance of greater objectivity. At the same time, since the evaluation is being conducted within the agency, there are evaluators who do know program strengths and weakness--needs and outcomes--and who can provide credible data and testimony to legislative bodies and funding sources. This is an attribute or advantage that outside contractors typically do not have in Washington.

The relationship between evaluating and agency policy making is less clear than that at the program level. The assistant superintendent heading that division which operates the programs approves the evaluation work plan, and the director of testing and evaluation briefly discusses the planned activities with the deputy superintendent and the SPI. These interactions, however, are frequently routine, resulting in statements like, "Sounds good, let's do it." The attitude is not negative or disinterested, rather it reflects confidence in the negotiated evaluation plan and the evaluation procedures being used. Simply stated, the position of the SPI policymakers seems to be: the job is getting done, there have been no great problems, funding sources are happy enough--why make changes? Why disrupt the process?

There are additional relationships. A computer playing a central role in the processing of evaluation data generates another set of interactions; and good working relationships between "man" and machine are crucial to the smooth implementation of the evaluation process. In 1976 a mini-computer was acquired to help solve the problems brought about by an abundance of work and a shortage of staff. Although the machinery and programming have become more sophisticated and the amount of data processed has grown, the basic human tasks have remained the same. Efforts go into the streamlining of application and reporting forms to make them more efficient for the entry of data into the computer and the aggregation of essential decision making information. The items on the forms are coded, if necessary, and entered into the machine. The reports of LEAs are printed out and returned to them for correction. The data are aggregated to match with the requests of progoram management, and the final updated computer files are used as the basis for preparing the annual evaluation report. Learning to work with the computer has been difficult for some program managers, but the system is expanding and providing a broader range of evaluation services each year.

USING EVALUATION

Definition and Purpose

In a recent roundtable discussion, members of the evaluation staff were asked to define "evaluation," and in every instance the responses represented a variation on two themes. First, evaluation was described as an objective process of collecting and organizing data, of judging impact and ascertaining value. Second, everyone agreed that the definition was not complete without a statement of purpose; for example, recognizing that evaluation is conducted to assist managers in making decisions.

Although this definition is broad enough to encompass the generally used strategies, two additional elements were described as necessary in a successful evaluation. These revolve first around the clear delineation of what data are required; that is, what questions are to be answered and, second, around the effective displaying and reporting of evaluative information.

Following this definition, the SPI evaluation effort includes: deciding what data are needed to answer the key evaluation questions, collecting and organizing the data, providing the necessary comparisons and judgments, and reporting the information in a way useful for decision making.

Ideally, the purpose of evaluation is to provide sufficient information about program alternatives so that managers can easily see the comprative value and make decisions that promote effectiveness and efficiency. The ideal situation generates information which reinforces the need for the program treatment and shows its impact. The evaluation activities ideally would follow a linear sequence in which needs were determined, objectives set, programs implemented, outcomes measured and information required to guide the next program cycle. In real life, however, the process is often abridged and seldom on the time schedule implied by the planning model. In most cases, the data generated are more important as formative information for program managers than as summary data for high-level policymakers, and certainly the data are more descriptive than judgmental.

Data collection is the easiest phase of evaluation, but it must be clearly established what decisions are going to be made, what data are needed, and when the analysis and report must be ready. In general, data are collected about the program targets, about the actual "performance" or outcomes and about the resources used. More specifically, several questions must be answered to guide data collection in the Washington evaluation process. What needs are addressed by the program? What objectives are included in the state plan? What implementation strategies are suggested? What outcomes are expected? What resources are provided? What groups are involved? Who is served? What are the treatments? How are the resources used? What are the program outcomes? What comparisons are appropriate?

Two additional questions influence the effort. Who is interested in the outcomes of the evaluation? Who ought to be interested? The answers are important for reporting purposes, but they also help to solve delineation problems in the future.

Five Uses of Data

1. Annual reports are prepared to meet the requirements of state and federal funding sources. These reports generally attempt to provide the answers to the previously mentioned questions guiding the data collection.

2. The data are also used to assist program managers in becoming more effective and efficient. Management memos are prepared as part of the annual reporting process, but the audience is SPI's program managers and policymakers, not the funding sources. The goal is to help the state to better meet its obligations through improved practices and quality control. The content of a memo may vary from the comments related to the need for improved office practices (for example, better written documentation of program changes), to the highlighting of objectives not met, or questionable fiscal practices. Although the program managers are not always pleased with the content, this use of evaluation data is viewed as a constructive practice.

3. The deputy and the superintendent rely on the evaluation section to keep them informed of any special circumstance that could ultimately require their attention. A third major use of the evaluation data is to make sure that policy makers are not suddenly confronted with an unpleasant surprise. They want to know in advance, for example, about the unpredicted concerns of special interest groups, anticipated major management problems, and possible audit exceptions.

4. An abundance of descriptive data is available in the computer files to provide information for decision making--a fourth major use. The files include, for example, numbers served, money spent, time in programs, delivery modes, staffing patterns, parent advisory committee activities, and program outcomes. The data are arranged by districts, by programs, and by funding source. The informtion is especially useful to program managers because it provides an up-to-date reflection on how resources are being used, that is, who is being served, in what ways, and at what costs. Through the aggregation of project

monitoring input, the managers can also review which projects are out of compliance and which rules or regulations are causing problems in the field.

5. Policy makers working at a different level of decision making abstraction use the data in two special ways. First, by reviewing the information available, they can keep the impact measures promised to funding sources as part of the accountability process more reasonable and within the range of possibility. This sensitivity is critical if credibility is to be established and maintained. Second, the data are used to support decisions previously made. Since often the timing of policy making and the collection and analysis of evaluation data cannot happen in the preferred sequence, the required data are estimated on the basis of past experience and updated when the actual data become available, hopefully to confirm the decision.

Evaluation and Policy Making

The term "policy maker" refers to the actions of the top-lelvel SPI managers as they develop the budget, the legislative thrust, and provide direction for the agency and the overall operation of the Washington common school system. This management level consists of the superintendent and his administrative staff, the deputy, and the five assistant superintendents who head the five agency divisions. The SPI is an elected official, and the members of this policy group serve at his pleasure and are exempt from the state civil service rules.

As specific policy questions arise, section directors, who are tenured state employees and provide professional continuity, are frequently invited to join in the policy-making proceedings. For example, the director of testing and evaluation provides significant input into policy decisions regarding the state level activities in that area. Describing this interaction as "providing input," however, does not reflect the full range of the dynamics. Although the process is not formalized in SPI operating procedures, there is an active two-way exchange. The director does participate in policy making related to testing and evaluation, but perhaps more importantly SPI policy makers rely on the section director to keep abreast of educational, legislative, and executive activities, both state and federal, and to take the initiative in providing them necessary information. The director, in a sense, is asked to be an advocate of sound professional practice and to also be able to discuss the impact that alternative decisions would have on various components of the educational community, with interests in testing and evaluation. In addition, it should be noted that just as open discussion and full input are expected before a decision is made, once it is made, everyone is expected to fully support its implementation.

The development of the state budget provides two examples of evaluation's involvement in policy making. Along with the amount of funds requested, each item in the budget must be defended with statements demonstrating the existence of need and describing the measures that will be used to show impact. The state testing program is a specific budget entry, so there is a direct policy making interaction regarding the activities planned for the biennium covered by the budget, the amount of money that will be required for implementation, and how the need and impact will be described in relation to the mandates of the state testing law. Since there is a tendency in organizations supported by budget allocations for sections or programs to attempt to show their value in terms of the amount of resources they command, the budgeting process frequently leads to a compromise. The demand for resources is greater than the supply, and a compromise between what is desirable and what is necessary by law and required for state leadership results. The series of negotiations is active and positive, and reflects both the superintendent's and legislature's priorities.

The testing and evaluation section is additionally involved in the budget making process through a technical assistance role. To maintain a solid reputation with the Governor's Office of Fiscal Management and the legislature, the needs, impact measures, and outcome data specified in the budget must be deliverable. Program managers and even the SPI executive staff are tempted occasionally to promise data that cannot be obtained. They are especially tempted to promise the measurement and reporting of achievement gain as an indicator of impact because this is the "hard" data that funding sources prefer. The problem is that in most instances it is not possible to deliver. Over the short run, the use of high sounding impact measures may bring funding; in the long-run, the loss of credibility outweighs the temporary advantage and causes major problems. The section is called on to help guard against this happening.

The development of the Washington Remediation Assistance Program provides another example of participation in policy making. In an effort to gain more resources for the support of the schools, the superintendent, with the advice of his administrative staff, decided to seek legislation and funding to promote the remediation of basic skills deficiencies in the intermediate grades. It was decided that results of the fourth grade testing program as the most "believable" data available would serve as the entitlement system for the allocation of the program's funds. The three years of background data were used to establish the need and the distribution mechanism, and the experience with Title I evaluation and the USOE models became a key element in the evaluation design. This latter connection was important, since the state remediation program was being organized to qualify for the federal incentive grants under Title I.

As mentioned in the discussion of the URRD program, the section also assisted in the development of RAP administrative guidelines and in the preparation of reporting forms. The achievement or impact section of the annual report will be prepared by testing and evaluation personnel.

AREAS OF FRUSTRATION

Like all evaluation units, the SPI's unit has faced many frustrations. Some situations have been solved, others circumvented, but a number, however, remain consistently unresolved and irritating. The following 12 situations provide brief examples of problems, both philosophical and practical, which continue to frustrate members of the Washington evaluation staff.

1. A clear role differentiation between evaluation and research has not been established. As a result, there is a range of instructional activities which have never been verified as efficient or effective, and there are probably a number of teachers who work very hard to do things which may not promote learning. Typically, the evaluator's prime role is to collect and organize data which describe program inputs and outcomes. In most instances, the researcher is interested in developing generalizations which explain or predict events. Neither researchers nor evaluators generally develop instructional materials or procedures, nor do they spend time checking to see if instructional methods are faulty or misused. There is a depressingly large twilight zone resulting from the unchecked assumption that the instructional methods used in a project accurately reflect the research findings and are being used appropriately. In fact, there is almost no research or evaluation energy applied to the analysis of alternative intervention strategies in Washington.

2. The role of "describing" in evaluation also requires clarification. The growing use of case study and ethnographical approaches suggest several questions: Where does the description stop and evaluation begin? How can the comparative statements so frequently demanded of evaluation be made? The potential of using these data collection techniques to assist in the program evaluation has not yet been fully realized, and the frustration of trying to harness the rich data into an evaluation statement with utility for an audience no on site persists.

3. Another frustration on the order of a "pet peeve" is the inability or unwillingness of program managers to separate monitoring from evaluation. Evaluation is clearly distinguished by an emphasis on program impact or outcomes. The on-site reviewing of projects to ascertain compliance with rules, regulations, stated objectives, and negotiated budgets is an important management function, but it is not evaluation. To consider that the worth of a project can be determined by the degree to which it is in compliance is misleading. The responsibility for monitoring as a management function is moving more and more to the program staff, and the energy and resources of the evaluation section are focusing on the evaluation questions of impact and efficiency. The movement is not complete; and to the extent that it is not, the frustration remains.

4. A fourth disappointment stems from the fact that too many educators are willing to use evaluation as an end in itself and to limit program emphases and alternatives to those amenable to "good" evaluation designs. Frequently, evaluators are blamed for causing this practice when they probably speak most loudly against allowing the evaluation to determine the program parameters. The situation has developed, or degenerated, in some instances, to the point where greater pride is taken in evaluation results than in the actual program outcomes, and this "Catch 22" scene appears to be growing.

5. A great frustration also results from aggregating impact and related data from 300 Washington school districts and watching important distinctions "wash out" in the averaging. There are successful projects and significant differences. However, implementing laws, for example, which call for the correlation of "appropriate input variables" with the achievement of grade four students tends to often obscure the situation rather than clarify it.

6. Using evaluation data inappropriately to respond to outside inquiry, or generalizing beyond the power of the data, is a persistent frustration. An example, once again from the fourth

grade testing, illustrates the problem. Frequently, letters come to the agency from people moving to Washington asking for help in deciding where to locate. The usual response is to send a copy of the fourth grade assessment report with district-by-district achievement results. This report shows fourth grade achievement scores, district level per pupil expenditures (a .08 correlation with achievement in 1976) and an average family income figure based on 1974 estimates (a .47 correlation with achievement in 1976). The sending of the report by implication suggests that it contains reasonable data for deciding where to live. One might say that this information is better than nothing, but the frustration is--it is not sufficient information for judging the quality of a school district.

7. Computer processing of evaluation data causes a frustration of major proportion, or perhaps more clearly, four frustrations. First, the battle is still being waged against the mentality that views the computer as something magic, with the ability to aggregate errors into precision. Second, the time-consuming problems of moving data from report forms into the machine are not fully appreciated by program managers, and the problems have not been solved completely. Third, working with a single, "one-owner" machine as compared to a service center, gives control, but a minor machine breakdown causes a major disruption of service. Fourth, the greatest frustration results when, even after lengthy planning and negotiations, program managers demand answers to questions that are not compatible with the data collected, stored, and programmed.

8. The problem of gaining sufficient support service in the preparation of evaluation reports has not been solved. The desired editing, graphics, lay out, and printing skills are not readily available in the agency, and going outside for assistance is difficult because of the rules regarding the role of the state printer.

9. The lack of congruence between the evaluator's "logic" and political decision making "logic" is also a keen source of frustration. The old and accepted political process of basing

decisions on power relationships is effective in lining up support and getting certain jobs done The fact that power politics works, however, does not detract from the frustration of presenting objective and overwhelmingly persuasive testimony based on accurate and logical evaluation data to a legislative committee and experiencing a contrary decision.

10. Another large source of frustration arises from the inability to gain a clear delineation from the policy makers regarding the decisions they will have to make, the information they will need, and when it will be needed. A major reason for the difficulty is that frequently evaluation data only play a marginal role in policy making. In the name of accountability programs must be evaluated, and the evaluation report must indicate what needs are being addressed and what outcomes are being obtained. Too often this information is treated as an end product, the report is made and filed, but the information is not used for planning purposes. Evaluators by default often carry on the delineation activities vicariously and hope for an accurate match with policy making needs. It is a difficult problem to solve, but repeatedly, clarification is lacking and the evaluation data are not on target and not useful for policy making.

11. An ultimate set of frustrations revolves around time. One problem is the nearly complete acceptance of the logical planning sequence which tends to mislead people into thinking that if they go through the steps they will automatically accomplish something. Evaluators move back and forth through the sequence in many different orders, and in most instances probably start by trying to establish what people with a stake in the program would be willing to accept as evidence that it is working.

A second time consideration that is frustrating could be labeled a continuity problem, and one specific example will elaborate the point. During a recent school year, representatives of the evaluation section joined with the university and educational service district staff to work closely with 12 small school districts on the Olympia Peninsula to help in planning and conducting assessments and evaluations aimed at clarifying priorities and isolating problem areas in the curriculum. The effort was productive, and by the end of the school year all involved had gained professional satisfaction for a job well done. The teams were eager to start the next year's round of activities. In September, however, the excitement faded--half of the districts had new superintendents and one district no longer existed.

A third time element is the recognition that time, or more specifically, timeliness itself, is a critical variable in evaluation. Perhaps the supreme frustration is to conduct a sound evaluation, generate useful information, and deliver a well documented report--just after the crucial decision has been made. In the spring of 1974, the legislature mandated that SPI conduct an evaluation consisting of pilot studies in LEA accountability and a statewide assessment of basic skills achievement. The law passed in April 1974, the funds became available July 1, and plans were developed in detail. Accountability projects were initiated involving a university, an educational service district, and several LEAs in the right geographic mix. An achievement test was developed using items from the National Assessment of Educational Progress, a scientific sample of students was drawn, and the testing scheduled for April 1975. All efforts were aimed at making best use of time and dollars and having the study's report completed by June 1975. Of course, the legislature met in January 1975, asked for the information, wondered what was taking so long, and nearly passed bad legislation.

12. A final frustration results from the fact that SPI staff and staff assignments are constantly changing. For example, over the past six years the Title I program staff has changed in some way every year, the management pattern for the migrant program has changed each of the past three years, and the assignments and personnel in the program for handicapped children have also changed dramatically. Federal legislation and reporting requirements in all of these areas have also undergone major transformation. In addition, reorganization within the agency

has become a way of life, and it seems as if the evaluation process always involves new people in one phase or another. Petronius Arbiter captured the idea when he reported his frustration in 60 AD:

> We trained hard, but it seemed that every time we were beginning to form up into teams we would be reorganized. I was to learn later in life that we tend to meet any new situation by reorganizing; and a wonderful method it can be for creating the illusion of progress while producing confusion, inefficiency, and demoralization.

Of course, Washington evaluators do not have a corner on this frustration.

IN CONCLUSION

During the 1970s, the emphasis on program evaluation gained widespread public popularity, as well as the strong support of executive policy makers and law making bodies. In fact, few developments have made so thorough an intrusion into the operating practices of education. There has always been an emphasis on the precision measuring and accounting for resources, such as the number of books in the library, pupil-teacher ratios, and the number of hot lunches served; but the stress placed on the evaluation of program results has come as an intricate part of the accountability movement. Other professions have concerned themselves with various forms of input-output analysis for many years; the decade of the 70s, however, marks the general introduction of accountability and program evaluation into the educational setting. The concern for the analysis of resources used and results gained is real and growing. However, the fact that there is a lot of "program evaluating" going on should not be interpreted as an indication of educational progress or confused with claims of program improvement.

Evaluation has come under criticism in recent times, especially from evaluators themselves. There has been a tendency, perhaps, to address too many of the tough problems related to the conduct of evaluation intellectually rather than practically--some of the most reputable evaluators are spending more time verbalizing about evaluation than practicing evaluation. One of the results is that new models, approaches, and strategies are developed and discussed, but the basic trouble spots plaguing "applied evaluation" remain unresolved. Crucial among these is the overestimation of the influence of evaluative information on management decisions and policy making. The intent of this paper was to provide background examples to use in thinking about evaluation issues and ways of improving practices.

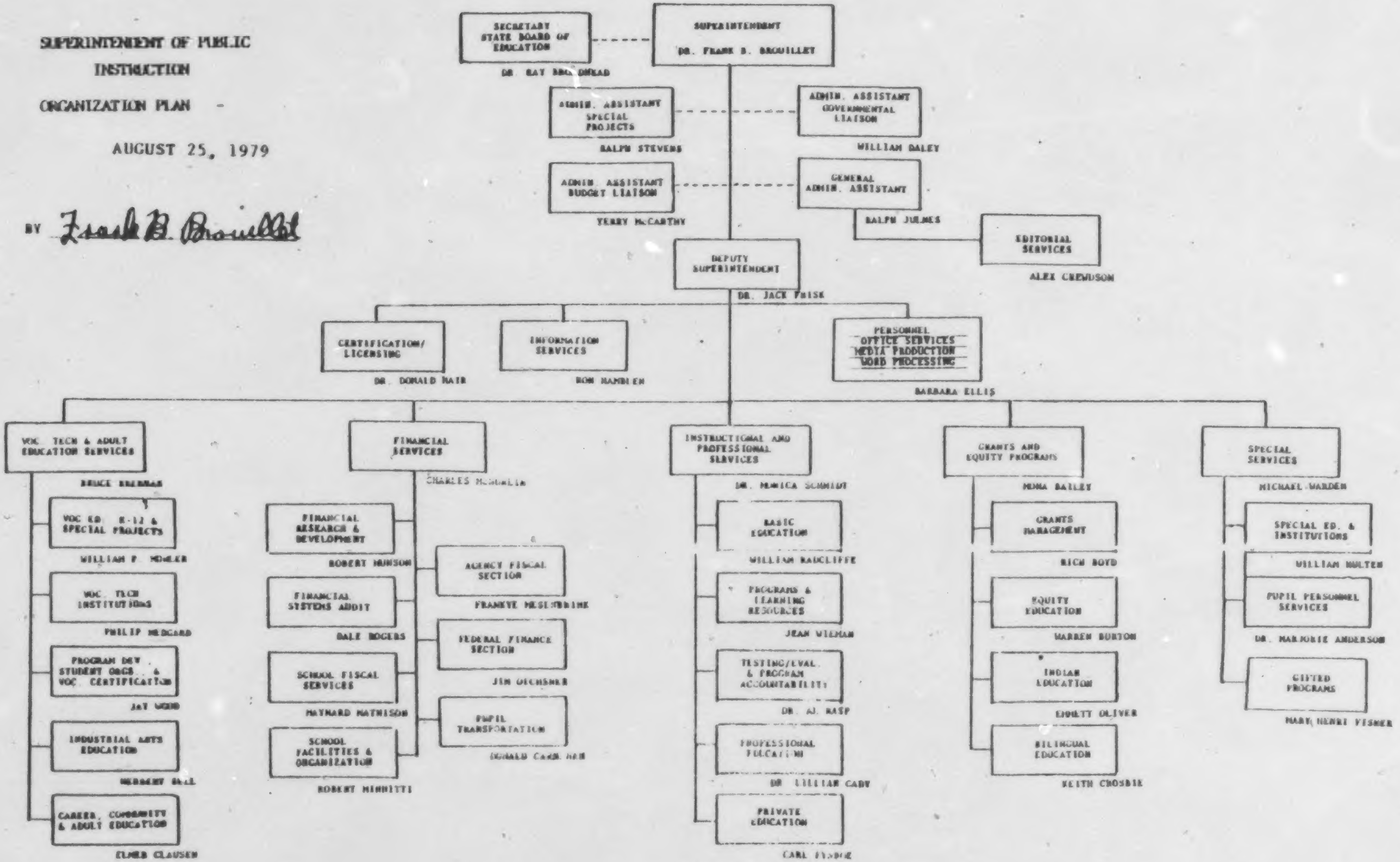
As stated earlier, few ideas have spread more rapidly to permeate the field of education than the concepts of accountability and program evaluation. Making the movement pay off with improved practices and better education for those served is still the challenge.

SUPERINTENDENT OF PUBLIC INSTRUCTION

ORGANIZATION PLAN -

AUGUST 25, 1979

BY Frank B. Brouillet



THE INTERACTION BETWEEN POLICY AND EVALUATION WITHIN THE SOUTH CAROLINA DEPARTMENT OF EDUCATION

Paul D. Sandifer, Director
Office of Research
South Carolina Department of Education

INTRODUCTION

Purpose

The purpose of this paper, as commissioned by the Research on Evaluation Program of the Northwest Regional Educational Laboratory, is to describe the interaction between policy and evaluation in the South Carolina Department of Education (SCDE). More specifically, how does policy influence evaluation and how, if at all, does evaluation affect policy within the agency?

A literal interpretation of the purpose leads one to attempt to view the SCDE in isolation from other state and federal agencies, as well as special interest groups. Although such an approach would have the advantage of resulting in a much briefer paper, it would ignore the considerable influence of other agencies and groups in shaping the policies of the SCDE and in establishing external policies under which the agency must operate. Consequently, the focus of the paper is not limited to the policy/evaluation interaction within the SCDE but also examines some of the ways in which various agencies and special interest groups affect policy and evaluation at the state and, consequently, at the local school district levels.

Although the paper focuses primarily on the SCDE and the interactions between it and other agencies, my perceptions of those interactions are no doubt colored not only by my particular responsibilities in the agency, but also as a result of seven years as an administrator in the Wyoming Department of Education and nine years as a teacher in public schools in Mississippi and Wyoming. An external observer, or other staff of the SCDE, might have perceptions of the policy/evaluation interaction that are quite different from my own. Since such differing perceptions are the rule rather than the exception, the writer recognizes that the thoughts expressed herein are but one version of the "truth".

Definitions

Although the terms "policy" and "evaluation" are widely used, and perhaps just as widely understood, it seems advisable to define the terms as they are used in the remainder of the paper. Policy, as used herein, is defined as including all legislation, regulations, position statements and policy statements, e.g., "the expressed policy of the State Board of Education is ---", the intended purpose of which is to determine a course of action, establish a program, or provide a framework within which decisions are to be made.

Evaluation, as used herein, is defined as the utilization of information, obtained through a systematic process of data collection, for any of the following purposes: assessing the impact of established policies or programs; comparing the effectiveness of two or more programs; assessing the degree of compliance with established policy; or influencing the establishment of new or revised policies or programs. This definition of evaluation deliberately avoids any attempt to draw the traditional academic distinctions between policy studies, research, evaluation, and assessment. This is done for two reasons, first, the common distinctions among these terms focus more on questions asked and procedures followed than on the use(s) made of the information collected, i.e., the distinctions are more semantic and academic than they are real and second, the results of research, evaluation (in the traditional sense), assessment and policy studies are all used, to varying degrees, in efforts to formulate or modify policies and programs. Whether a particular data collection effort should legitimately be labeled an evaluation, seems to be more appropriately determined by the use(s) made of the data than by the particular study design or the procedures used in collecting the data. Regardless of the complexity and/or degree of sophistication of an "evaluation" design, the act of collecting data does not constitute evaluation. Evaluation occurs only after the data are collected and then, only if the data are used as a basis for making judgements about worth, value or effectiveness. Although they may not be acted upon by those in policy setting positions, the first place such judgements are normally identifiable is in the evaluation report.

Although they are consistent with the definition given here, many of the examples of evaluation used in this paper will not be regarded by academicians as "true" evaluation. However, the broad definition of evaluation previously given is necessary in order to understand the policy/evaluation interaction.

Organization of Paper

The remainder of the paper is comprised of four major sections. The first, "South Carolina Public Education: State Organization and Administration", provides a description of the context within which the writer's perceptions of the policy/evaluation interaction have been formulated. The second, "Policy: Influence on Evaluation", concerns the ways in which policy determines what is to be evaluated and the impact that policy has on evaluation methodology. The third, "Evaluation: Influence on Policy", represents the writer's perception of the conditions under which evaluation does, and does not, influence policy. The final section, "Other Factors Influencing Policy and Evaluation", examines the impact which special interest groups have on the formulation of policy, the design of evaluation, and the uses to which evaluation findings are put.

SOUTH CAROLINA PUBLIC EDUCATION: STATE ORGANIZATION AND ADMINISTRATION

State Superintendent of Education

The Office of State Superintendent of Education was established by the Constitution of 1868. The Superintendent is elected on a partisan ballot for a four year term and there is no limit on the number of consecutive terms which the Superintendent may serve. During the period from reconstruction until 1979, twelve superintendents were elected to office. The current Superintendent, Dr. Charlie G. Williams, began his first term in January, 1979.

The general duties of the State Superintendent, as prescribed in the School Laws of South Carolina (1976), include:

1) serving as secretary and administrative officer to the State Board of Education;

2) supervising and managing all public school funds provided by the State and Federal Governments;

3) organizing, staffing, and administering a State Department of Education; and

4) administering, through the State Department of Education, all policies and procedures adopted by the State Board of Education.

State Board of Education

The State Board of Education is comprised of seventeen members, one from each of the sixteen judicial circuits and one member at large. The members from judicial circuits are elected by the legislative delegations representing the counties of each circuit. The "at large" member is appointed by the Governor. The terms of the members are four years and no member may serve consecutive terms except by the unanimous consent of all members of the county legislative delegations from his/her judicial ciruit. The statute pertaining to the composition of the State Board contains no provisions excluding professional educators from service on the Board. The present chairman, and five other members are professional educators. Although the members are elected by their legislative delegations, the practice of electing educators has received criticism from some members of the General Assembly. During the past several years, legislation has been introduced, but not enacted, to restrict State Board membership to the lay public.

The general powers of the Board include:

1) adopting policies, rules and regulations not inconsistent with the laws of the State for its own government and for the government of the public schools;

2) annually approving budget requests for the institutions, agencies and services under the control of the Board;

3) adopting minimum standards, for any phase of education, as are considered necessary to aid in providing adequate educational opportunities and facilities;

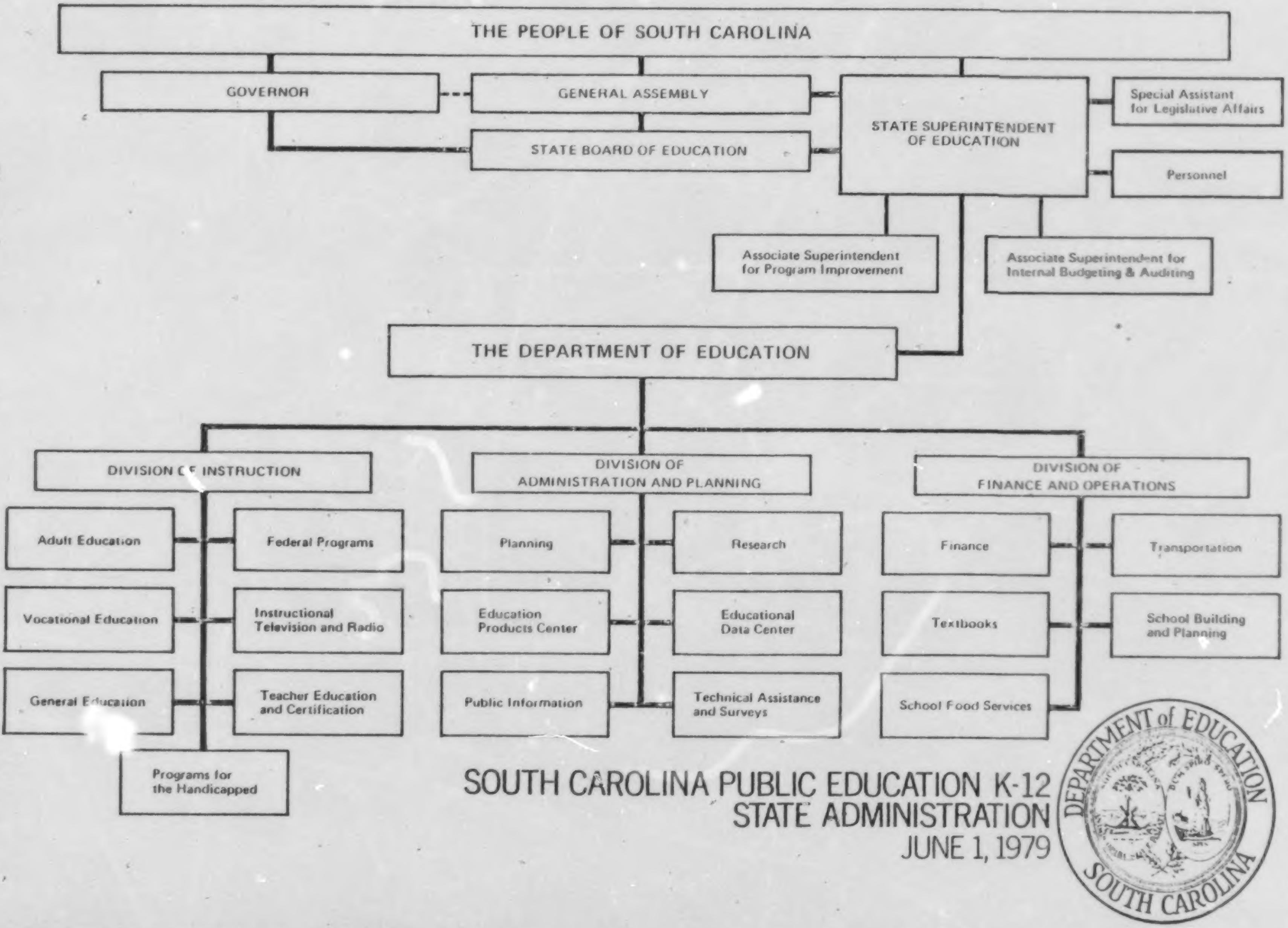
4) prescribing and enforcing rules for the examination and certification of teachers; and

5) prescribing and enforcing courses of study for the public schools.

State Department of Education

The administrative structure of the State Department of Education includes three divisions which are under the supervision of Deputy Superintendents who are directly responsible to the State Superintendent of Education. Each of the three divisions, Administration and Planning, Instruction, and Finance and Operations, includes several offices which, collectively, administer the programs for which the agency is responsible. Although most of the offices include two or more sections, the organizational chart (Figure 1) does not include detail below the office level. The organizational pattern of the Department has remained relatively stable during the five years in which the writer has been an employee. The only significant changes, the creation of the positions for associate superintendents and special assistant for legislative affairs, as well as placing the Office of Personnel under the direct supervision of the State Superintendent, have

Figure 1


THE PEOPLE OF SOUTH CAROLINA
GOVERNOR
GENERAL ASSEMBLY
STATE BOARD OF EDUCATION
STATE SUPERINTENDENT OF EDUCATION
Special Assistant for Legislative Affairs
Personnel
Associate Superintendent for Program Improvement
Associate Superintendent for Internal Budgeting & Auditing
THE DEPARTMENT OF EDUCATION
DIVISION OF INSTRUCTION
DIVISION OF ADMINISTRATION AND PLANNING
DIVISION OF FINANCE AND OPERATIONS
Adult Education
Federal Programs
Vocational Education
Instructional Television and Radio
General Education
Teacher Education and Certification
Programs for the Handicapped
Planning
Research
Education Products Center
Educational Data Center
Public Information
Technical Assistance and Surveys
Finance
Transportation
Textbooks
School Building and Planning
School Food Services

SOUTH CAROLINA PUBLIC EDUCATION K-12
STATE ADMINISTRATION
JUNE 1, 1979

occurred since January, 1979.[1]

The Department employs 1079 individuals of whom approximately one-half are involved in the maintenance and operation of the state supported pupil transportation system. With the exception of the employees of the Office of Transportation, most of the staff are based in the agency offices in Columbia.

Evaluation functions within the agency are decentralized. Although no office title within the agency includes the word evaluation, several offices carry out activities which fall within the broad definition of evaluation which was presented earlier. There is, however, no State Board or agency-wide policy concerning evaluation responsibilities.

Each office which funds programs operated by school districts has, or assumes, the responsibility for evaluating, or monitoring the evaluation of, all programs which it administers. The determination of which programs are actually evaluated is, more often than not, a function of federal mandates. The offices most heavily impacted by federal mandates for evaluation are Federal Programs, Adult Education, Vocational Education, and Programs for the Handicapped.

Two offices which do not administer funds for locally operated programs but which are involved in evaluation activities are Technical Assistance and Surveys, and Research. The Office of Technical Assistance and Surveys conducts, at the request of local school boards, studies to determine needs in the areas of administration, curriculum, personnel, and facilities. The results of these studies frequently provide the basis for district planning to meet the identified needs. Although in the traditional sense these studies might not be considered evaluations, they do provide information on which policy is based and actions are taken.

The office of Research is involved in evaluation activities in three ways. The first, and most time consuming, is through the administration of the Statewide Testing and Basic Skills Assessment

---

[1] The State Superintendent recently initiated a management review of the agency for the purpose of determining whether some organizational changes may be desirable in order to more effectively and efficiently fulfill the responsibilities of leadership, service, and regulation. As a result of the study, the organizational chart (Figure 1) may be inaccurate by the time this article appears in print. With that possibility in mind, I have attempted to keep my observations specific to the agency (as requested) while at the same time keeping the interpretations general.

Programs. The second involves data collection to assess the financial impact of proposed or existing policies, e.g., "What is the projected cost for facilities required to implement a legislative mandate to reduce pupil/teacher ratios in grades one through three?". Finally, evaluations are occassionally conducted at the request of other offices within the agency or as a result of a decision of the State Superintendent.

POLICY: INFLUENCE ON EVALUATION

Organizational Influence

Whether the decentralization of evaluation responsibilities in the SCDE is more a function of default than conscious decision making is not known. However, evaluation as a recognized responsibility of State Education Agencies is still in its childhood, or at best early adolescence, and some agencies have not chosen to organize in a manner that concentrates that responsibility in one unit. Experience gained by serving in two state education agencies, one in which the evaluation function is centralized and one in which it is decentralized, indicates that both organizational patterns have their unique disadvantages.

The influence of agency structure on evaluation is evidenced in several ways. When the responsibilities for evaluation are decentralized there may be no common criteria which are uniformly applied either in the employment of staff or in the design of evaluations. This frequently results in:

1) considerable variation in the level of expertise of evaluation staff assigned to various offices within the agency;

2) a greater than acceptable degree of variation in the quality of evaluations; and

3) a lack of consistency in the kinds of evaluation requirements which the various offices impose on local school districts.

When evaluation responsibilities are decentralized, evaluators are frequently directly responsible to the administrators of the programs for which they have evaluative responsibility. Even if objectivity can be maintained in such situations, evaluation findings may lack credibility because of an apparent conflict of interest. This situation, however, is not automatically overcome by centralizing the evaluation function within the agency. Individuals or groups external to the agency may still consider findings with which they disagree to be merely a reflection of the agency's bias.

With centralization of the evaluation function, the disadvantages cited above are eliminated or at least alleviated. On the other hand, if evaluators are assigned to a unit that has no programmatic responsibility, they may well be viewed with suspicion and distrust by the administrators of those programs which are to be evaluated. Additionally, communications across or through administrative channels in a bureaucracy can be both slow and frustrating.

Whether responsibilities for evaluation are centralized or decentralized is probably less important than having an agency commitment to "good" evaluation for the purpose of addressing policy relevant issues. In the absence of such a commitment, it is unlikely that agencies will anticipate information needs and consequently will frequently be placed in a reactive rather than proactive role in the policy making process.

Federal Influence

Policy, regardless of the governmental level at which it is created, apparently influences evaluation in two major ways: first by determining what is to be evaluated, and second, by shaping or dictating the evaluation methodology.

Determining what is to be evaluated is, in many instances, not a matter of choice for any State Department of Education since State and Federal statutes and regulations, e.g., ESEA Title I, may be very explicit in this regard. A major portion of the evaluation efforts of the SCDE are directed toward complying with federal mandates although federal funds account for only approximatley 14% of the annual expenditures for public elementary and secondary education in South Carolina. This is not to say that too many resources are expended in evaluating the effectiveness of federally funded programs. It does suggest, however, that in the past we have probably spent far too little time and money in evaluating state funded programs.

The impact that policy has on shaping or determining evaluation methodology is nowhere more easily identifiable than in the evaluation requirements for ESEA Title I. The federal regulations stipulate the only evaluation methodologies that may be used by states and local school districts. Any exceptions to the prescribed methodologies must be approved by the U.S. Commissioner of Education. The expressed rationale for these models is that they will yield comparable data on pupil achievement that can be aggregated to the national level, i.e., across school districts and states. Various critics have raised questions about the validity of this assertion because of some unresolved technical issues surrounding the models. Assuming, however, that the models can and do yield reliable and valid data, another basic question is still unanswered. How can aggregated pupil achievement data be used in addressing a policy issue which, on the face of it, is more social and political than educational in nature? Would it not be sufficient, and perhaps more appropriate, to determine: 1) whether the target population as defined in legislation is actually being served; and 2) what type of instructional programs or what organizational patterns are most effective in meeting the needs of the educationally disadvantaged?

As a member of the Evaluation Sub-Committee of the Committee on Evaluation and Information Systems (CEIS) of the Council of Chief State School Officers, the writer has been priviledged to hear much of the debate concerning the Title I models. Apart from the questions which have been raised about the technical quality of the models and the policy relevance of the data, concerns have also been voiced about the "test only" approach to evaluation. Local district personnel are concerned that the use of Title I funds for evaluation will be restricted to the collection of data required by federal regulations and consequently, their efforts to examine the effectiveness of other program components will be severely hampered.

In addition to the Title I evaluation models which are already being used, the U.S.O.E. is developing models for the evaluation of programs for migrants and for children in institutions for the neglected and delinquent (N and D). In the first draft stage, the models for programs for N and D are also pupil achievement based. The methodology to be used is certainly being determined by policy which will probably be promulgated in the form of regulations.

Beyond the impact that specific program policy, e.g., Title I, may have on evaluation methodology as applied to that program, the effects often carry over into other programs. For example, the comparability and non-supplanting requirements of Title I, coupled with the Office of Civil Rights regulations which prohibit grouping that results in the formation of racially identifiable classes, virtually prohibits the use of experimental or quasi-experimental design in evaluating programs that may have little, if any, relationship to the federal programs which have placed constraints on evaluation in general.

State Influence

Three recently enacted South Carolina statutes pertaining to education include explicit evaluation requirements. These are the "South Carolina Education Finance Act of 1977"; the "Basic Skills Assessment Act of 1978", and the "Teacher Training, Evaluation and Certification Act of 1979".

The Education Finance Act includes an "accountability" section which requires:

1) the establishment of school advisory councils;
2) school and district based needs assessments;
3) the development of annual plans to meet identified needs;
4) district participation in the statewide testing program as prescribed by the State Board of Education; and
5) annual reporting of program effectiveness to the general public and the State Board.

The Basic Skills Assessment Act, although not as explicit in its requirements as are the federal regulations concerning Title I, contains requirements which effectively shape methodology and instrumentation. The law requires:

1) the administration of a readiness test at the beginning of grade one;
2) tests of reading and mathematics at the end of grades one, two, and three;
3) tests of reading, mathematics, and writing at the end of grades 6 and 8; and
4) a test of "Adult Functional Competency" at the end of Grade 11.

The law includes a time-table for the development and implementation of the program and stipulates that the tests shall be criterion-referenced. Further, the tests shall be used for the purpose of diagnosing student deficiencies and as a basis for remediation. The tests are not to be used as a basis for promotion or non-promotion. None of these stipulations are necessarily undesirable, but they do have considerable impact in shaping the assessment program.

The Teacher Training, Evaluation, and Certification Act mandates major changes in teacher training and certification procedures. The requirements of the legislation include:

1) all applicants for admission to teacher education programs in State supported institutions must successfully complete a basic skills examination in reading, writing, and mathematics;

2) the development of an instrument to be used by colleges and universities in evaluating all student teachers;

3) development of an instrument to be used by local school districts in evaluating teachers during their provisional year of certification;

4) successful completion of a teaching area examination as one requirement for provisional certification; and

5) discontinuing the use, after July 1, 1981, of the Commons Examination of the National Teacher Examinations for the purpose of teacher certification.

The legislation includes a number of other provisions but those cited appear to be the ones which impact most significantly on evaluation procedures and methodology.

Not only does policy influence evaluation by determining what is to be evaluated and what methodology may be used, it also has impact on local acceptance of programs. Local district personnel frequently perceive evaluation requirements as being somewhat arbitrary and infringing upon their rights to make decisions locally regarding curriculum and instruction. Questions that are of interest to funding agencies (state and federal) may be of little, if any, interest to local district administrators. For example, few district administrators are interested in or concerned with the external validity of a project or program which is operating within their school district. Whether the project or program works in their particular district is of more concern than whether it may be exportable to some other district similarly situated. Evaluation design is, however, determined by the questions of interest to the funding agency and external validity is of interest if the agency is considering replication of the project. Although the funding agencies may ask legitimate questions, lack of sensitivity to local needs does little to gain the support and cooperation of local district personnel who are implementing the program.

All evaluations are not initiated as a direct result of statutes, rules, regulations, or other written policy statements which may require that evaluations be conducted. Occassionally, evaluations are requested long after a course of action has been determined and a program has been implemented. Such evaluations are usually sought as a means of either providing data to generate continued support for a decision that was initially based primarily on beliefs or for generating support for the expansion of the program. This always raises the question of whether what is being sought is really an evaluation or a "Good-Housekeeping seal of approval". There is no intended implication that those who seek such "legitimatizing" evaluations are dishonest or unethical. To the contrary, they are almost always sincere, dedicated individuals who firmly believe that their program is working, has great merit, and should either be maintained or expanded. Seldom have these "stake-holders" entertained the possibility prior to conducting an evaluation, that the results may not support their biases.

Although the questions addressed by such "legitimatizing" evaluations are not directly influenced by the policies or actions that established the programs being evaluated, the methodology is certainly influenced by their _ex post facto_ nature.

EVALUATION: INFLUENCE ON POLICY

A common lament of evaluators is that the results of their efforts are not used by decision makers. Although this is not always the case, the situation occurs frequently enough to cause great concern among those who practice the art (or science) of evaluation. Assuming that decision makers are reasonably rational individuals who would prefer to make decisions on the basis of information rather than intuition, there must be reasons why evaluation findings are not used. Experience indicates several possibilities:

1) the conclusions of the evaluator are not germane to the decisions that must be made;

2) the evaluation findings are not reported in a manner that communicates to the policy makers;

3) the findings do not support the biases of those making policy decisions; or

4) a lack of credibility results either from poorly conceived or conducted evaluation or the reporting agency not being credible to the potential users of the data.

Each of these possible reasons for ignoring the results of evaluation is expanded to some degree in the remainder of this section.

Conclusions Not Germane

Educational evaluators tend to function in the manner that the label implies, i.e., given freedom in designing evaluations they tend to focus only on the educational and cost aspects of program or policy and usually ignore, or fail to realize, that there are often social and political, as well as educational and financial issues involved. The evaluation of any program or policy may well be of interest to more than one group of decision makers and these various groups may legitimately be interested in answers to different questions. For example, the local district administrator of a federally funded program may be interested only in whether the program produces gains in student achievement beyond that which would be predicted without program intervention. The state level administrator's interest may extend to the question of external validity. Administrators at the federal level may share the interests of state and local administrators, but in addition they want data that can be aggregated across states, whereas, the funding body, Congress may be more interested in the social and political aspects of the program, e.g., is the group for which the program was intended actually being served? Compensatory education programs are a case in point. There is serious doubt that Congress will ignore social and political issues and make significant changes in any compensatory education programs solely on the basis of pupil achievement data as a measure of program effectiveness.

Evaluations which are too narrowly focused yield results which may be viewed as not germane to the issue to be decided. An example is available as a result of an action by the South Carolina General Assembly which provided funding for a pilot program to reduce pupil/teacher ratio in the first grade from 26:1 to 20:1. The Office of Research in the SCDE was assigned the responsibility for evaluating the pilot program. Since the intent of the pilot program was "obviously" to determine the cost-effectiveness of a reduction in pupil/teacher ratio as a means of increasing student achievement, an evaluation focused on achievement gain was designed and conducted over a two-year period. The results of the evaluation were consistent with a large body of the literature on the subject and indicated that a reduction in pupil/teacher ratio of the magnitude involved was not a cost effective means of increasing student achievement[2]. Shortly after the evaluation results were released, the General Assembly enacted the Education Finance Act of 1977 and mandated a reduction of pupil/teacher ratio from 26:1 to 20:1 in grades one through three. Why? In retrospect, the evaluation was too narrowly focused and failed to take into consideration the political aspects of the issue. The South Carolina Education Association, a rather effective lobbying group, was exerting pressure on the legislature and as a means of gaining their support for the total finance bill, the proviso relating to pupil/teacher ratio was included. The question addressed in the evaluation certainly seems to be a legitimate one but perhaps the utility of the study could have been improved by including a survey of teachers to determine whether they preferred to have the available funds used to reduce pupil/teacher ratio or to provide an increase in salary. Additionally, taxpayers, especially those who are parents of school-age children, may be supportive of smaller classes because of their beliefs that children receive more individual attention in smaller classes. The point is, there were other questions, in addition to those related to gain scores, that probably would have been of interest to the policy-makers and, consequently, should have been included in the evaluation.

---

[2] Among the common problems encountered by evaluators are those created by insufficient time and financing. If it had been possible to conduct a longitudinal study of the students involved in the program, the results might have been different. Although the educational system is generally concerned with long-term benefits, evaluation is frequently restricted to the examination of immediate outcomes.

If we are really interested in increasing the use of evaluation findings in determining policy, more attention must be given to identifying the social and political, along with the educational and economic, issues involved. No one is so naive as to believe that all decisions affecting education are made by educators, however, the narrow focus of many evaluation designs is not consistent with our knowledge of the decision making process. The answer to the question "By whom will the results be used?" is too often left until the final report is being written, in which case, a likely answer is "no one". If evaluations are to yield policy relevant results, the various stake-holders and the questions of interest to them must be identified before the fact.

Evaluation Reports That Do Not Communicate

The policy maker who has the technical background requisite to the interpretation of the typical evaluation report is about as rare a specimen as a 5'2" professional basketball player. All too frequently, evaluators appear to be more interested in producing reports for journals or discussions with their colleagues than in communicating results to potential users of the data. The development of highly technical reports may work wonders for the evaluator's ego but the intended users may be left with two choices: 1) ignore the report, or 2) request assistance in interpreting it. A colleague once made a learned, but incomprehensible, presentation to a lay advisory committee. Following his one-hour monologue, during which the committee members were attentive and polite, two committee members made comments. The first, a newspaper editor, said, "I consider myself to be a reasonably intelligent man, but for the past hour you have insulted my intelligence by subjecting me to your jargon." The second, a wealthy rancher, commented, "As a former school board member, I made it a practice never to fund anything that I could not understand." Unfortunately, the message the committee members were trying to convey was lost on the speaker. The point is, that as long as we evaluators focus our communication efforts primarily on each other and the academic community, we have little room for complaint when the results of our efforts are not used by decision makers.

The Office of Research in the SCDE invested considerable time and effort in determining correlates of achievement test scores in order to generate predicted school mean achievement scores for comparison with those obtained through the Statewide Testing Program. The intended use was to stimulate local school administrators to take a closer look at schools in which there were significant differences between the predicted and obtained scores. The Office prepared and disseminated to district superintendents a "non-technical" summary report on the project. One recipient of the report wrote to the State Superintendent of Education and commented that if he was really expected to understand the report, it was probably a good thing that he had only a short time left prior to retirement. For those of us involved in preparing the report, the easiest thing to do was to agree that it was probably good that his time of retirement was near. In reality, however, he was probably not the only one of the ninety-two superintendents with whom we failed to communicate. The reports have since been revised and are now being used as initially intended in many of the South Carolina school districts. Unfortunately, we do not know the extent to which the potential utility of the information was reduced by our initial failure to communicate.

Results That Do Not Support Stake-Holder Bias

Precedent to a decision to evaluate the impact of any program or policy is the decision that created the program or formulated the policy. This decision usually reflects some belief of the policy makers that the course of action being pursued is appropriate and therefore should yield beneficial results. In essence, the stake-holders have evaluated, with positive results, the course of action before pursuing it. Consequently when formal evaluations result in findings contrary to the policy makers' bias, the evaluator, and his/her findings, may encounter considerable resistance. The ego involvement of a policy maker may be so great that either revising an existing policy or formulating a new policy contrary to the previousy selected position represents a cost which is too great to be paid. In this case the policy maker may simply disregard the evaluation results and proceed on the previously established course. On the other hand, the initial rejection of the findings may, with time, give way to a change in position. The writer has observed this shift occur so gradually that it was virtually impossible to determine the point at which the change was made. (This seems to have implications for studies of evaluation utilization. In what time-frame must the results be used in order to consider them "utilized"?).

For several years the SCDE supported the implementation of extended-day kindergarten programs in the belief that such programs were efficacious in developing "readiness" for first grade. When the results of a third-party evaluation indicated non-significant differences between half-day and extended-day programs in this regard, the findings met with considerable resistance and unjustifiable questions were raised about the technical quality of the study. Gradually, however, the program administrators have shifted their position on the issue. The point to be made is that it may be unreasonable to expect negative evaluation findings to be immmediately and warmly embraced by those who are stake-holders in the program. Perhaps the best that can be hoped for is that when decision makers are given both information and time, reason will prevail.

Lack of Credibility

As stated earlier, non-utilization of evaluation findings due to lack of credibility may possibly be attributed to one of two reasons: 1) the evaluation was poorly conceived and/or conducted, or 2) the agency reporting the results lacks credibility with the potential users of the data. The only apparent defense against the first of these is to conduct other evaluations and not make the same mistake twice. Unless, however, the services of a different evaluator are utilized, the results may still be viewed with suspicion by the target audience.

In the second instance (lack of credibility) there, again, is little that can be done after the fact. Prior to engaging in any evaluation effort, the agency should carefully consider whether the potential users of the data are likely to consider the results to be objective and unbiased. When the evaluation function is centralized within the Department of Education and, consequently, evaluators are not answerable to program administrators, evaluation results may have credibility within the agency itself. On the other hand, when viewed externally, even a centralized evaluation function is still a part of the total agency and results may be suspect. Unless, the results of the evaluation are solely for internal decision making, this probably argues against having any State Department of Education evaluate programs which it administers.

Other Factors Influencing Policy and Evaluation

Although the influence which special interest groups have upon the formulation of policy and, consequently, upon evaluation has been mentioned earlier, the magnitude of this influence is such that it seems to warrant some special attention. Policy is not formulated in a vacuum. Policy is formulated only in response to external influence or, more to the point, in response to some real or imagined need. In the opinion of the writer, any discussion of policy/evaluation interaction which does not consider those external influences ignores the force which drives the system. Regardless of whether a policy is developed by the Congress, the State Legislature, the State Board of Education, or the State Superintendent, it appears that the policy is always in response to the stated, or inferred needs or desires, of some group. For example, members of the legislature, as duly elected representatives of the public, enact statutes (educational and otherwise) which, at least in theory, reflect the desires and interests of the electors. The State Board of Education then translates the education related statutes into rules and regulations which are administered by the State Superintendent of Education. In turn, local boards of education may develop their own policies designed to implement those imposed upon them.

Any of the policy-making groups, the legislature, the State Board of Education, or local school boards may generate policy in response to their own constituents without being motivated to do so by one of the higher level policy-making bodies. In any case, however, the link between external influences and policy seems obvious.

This translates, then to a system of interaction that might be portrayed as in Figure 2. The implications are that special interests directly influence policy and, indirectly, evaluation; policy directly influences evaluation; and evaluation _may_ influence policy either directly by feedback to the policy-makers or indirectly through feedback to special interests.

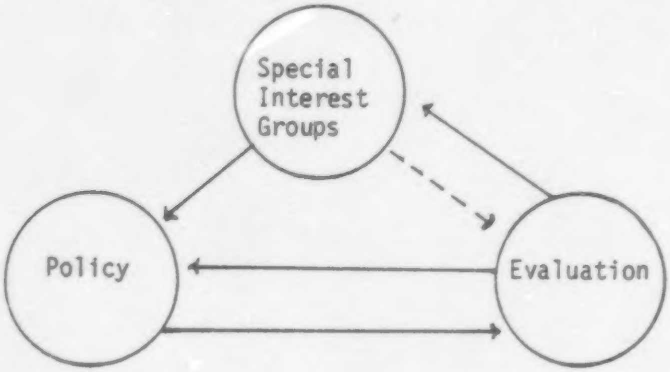


Figure 2

Unfortunately, the process is not nearly so straightforward as the preceding paragraphs and Figure 2 may imply. There seem to be at least two sources of "noise" in the system that have implications for evaluation and the use of evaluation findings. First, there are usually competing interests at work any time that policy (law, rule, or regulation) is established. Consequently, compromise may be made in framing the policy with the result that no one is completely satisfied with the outcome. If this is true, various "stake-holders" may seek quite different results from evaluations or at least may have different biases toward those results. Second, even if the policy is a precise statement of the expressed desires of the interest group, the underlying intent may not be reflected in the policy. This may lead evaluators down the primrose path to asking the wrong questions (or at least not enough questions) if they mistakenly assume that a policy is addressing an educational issue when, in fact, the force behind the policy was more of a social, political, or economic nature. The earlier reference to the question of reducing pupil/teacher ratio provides an example of the pitfalls which can be encountered as a result of either ignoring, or failing to recognize the motivation for a particular policy.

In concluding, it seems desirable to provide an example of the special interest/policy/evaluation interaction. In using the South Carolina Basic Skills Assessment Act of 1978 as an example, the writer is aware that his perceptions constitute only one version of the "truth".

The legislative history of the act began in 1977 when two representatives, both members of the Black Caucus, co-sponsored a bill calling for the implementation of a program of grade by grade promotion based on achievement test scores. The bill provided that the State Department of Education would select the tests and determine the required scores. The proposed program was to be implemented over a twelve year period beginning with grade one and adding a grade each year until all grades were included. The bill did not receive favorable action in the General Assembly but it did generate sufficient interest to result in a joint resolution creating a special committee to study the issue of minimum competency in the basic skills.

In its report to the General Assembly, the committee included a draft of proposed legislation. Although over eighty amendments were introduced during the passage of the bill in 1978, the law is essentially the same as the bill proposed by the committee. There is, however, one major difference. The committee's proposal that an eleventh grade test of "Adult Functional Competency" be required for a high school diploma was deleted prior to passage of the act. What began in 1977 as a move to require grade by grade promotion based on testing, culminated in 1978 with a required testing program to be used for the diagnosis of student deficiencies and as a basis for providing basic instruction to assist students in overcoming those deficiencies. Major compromises, the result of the interests of various groups, were reached along the way.

What motivated the introduction of the original bill in 1977? The most obvious answer is that it was simply an extension of the national trend towards competency testing. The prime sponsor of the bill says that such was not the case. According to the sponsor there were two factors which prompted him to introduce the legislation. First, his constituents were concerned that many of their children were only semi-literate upon graduation from high school and they were seeking a remedy for that situation. Second, the legislator had access to data collected through the SCDE operated Statewide Testing Program which indicated that, on the average, the achievement scores of minority students were considerably below those of the majority. He attributed these differences to poor teaching and social promotion. He apparently perceived promotion as a function of testing to be a remedy for the situation.

The Basic Skills Assessment legislation is a rather significant policy statement concerning evaluation. Apparently it came into being as a result of the influence of special interests and the use of available evaluation data that supported the concern of the interest group.

Conclusion

This paper was written for the purpose of providing a non-academic perspective of the policy/evaluation interaction in a state education agency. While I am confident (at the 99% level) that it is non-academic, I am not equally confident that it provides a "true" picture of the interaction which was the subject of the paper. As stated earlier, others may, and probably do, have quite different perceptions of the ways in which policy and evaluation interact, or fail to do so, within the SCDE. But that is the nature of evaluation, take the information and make your own judgement.

References

Act No. 187 of the 1979 Acts and Joint Resolutions of the S.C. General Assembly.

Code of Laws of South Carolina, 1976, Chapters 3, 5, 20, and 30.

Report of the Special Joint Education Committee to Study Minimal Competency in Basic Skills, 1977. The Special Joint Education Committee to Study Minimal Competency in basic skills created by Part II, Section 31, of Act 219 of 1977.

POLICY AND EVALUATION:
A CONCEPTUAL STUDY*

Thomas F. Green
Syracuse University

What is (or can be) the relation between public policy, on the one hand, and evaluation, on the other? Is there a way to attain clarity in understanding what an appropriate relationship would be? I intend to answer these questions in two steps. I shall consider first the typical character of policy questions. In a second part, I shall examine different facets of the policy process--analysis, formation, decision, and political judgment. I believe that these steps, taken together, will allow us to form a view, however tentative, about the relevance of evaluation as a professional practice to various aspects of public policy.

---

*The numbered footnotes in this paper are not really footnotes. They refer to corresponding items in the attached Appendix. This paper can be read as an independent analysis. It constitutes an attempt to reach conceptual clarity on certain features of the relation and interrelation between the practice of evaluation, on the one hand, and the formation of public policy, on the other.

But the numbered items in the Appendix establish specific links between this paper and a body of six others commissioned by the Northwest Regional Educational Laboratory dealing with the practice of evaluation in State Education Departments.

The numbered items, then, are citations of points in those other papers that either provide the practical illustration of points made in the text or have actually provoked points made in the text. The underlying principle is that a philosophical analysis of such matters should, on the one hand, illuminate what practitioners do and say about their work; but, on the other hand, it should also arise from a serious study of what they say and do. Philosophy, in this sense, is simply the explication of everyday life.

I

Policy Questions

There is probably no single definition of "policy" adequate to capture the full range of ordinary usage. Such a definition would have to satisfactorily capture the likenesses and differences between managerial decisions, guides to practice, rules of legislation, basic choices of political direction, and the bar application of standard requirements in administration (No one gets unemployment assistance for more than twenty-one weeks). It would have to include some matters that fall under "Standard Operating Procedures" (file expense accounts within ten days with receipts), matters of personal practice (I don't answer the phone at home because it's never for me. Avoid arguments in the office. Don't give G-3s unsatisfactory ratings; it takes too long to defend them.) Although each of these things can be called "policy", the term has, in each case, a slightly different meaning.[1]

On the other hand, it is possible to establish some common features of what we ordinarily take to constitute a policy question, especially if our concern is with public policy. So I wish to render the question, "What is policy?" by asking "What is a policy question?"

A policy question is a request for a line of action aimed at securing an optimal resolution of a conflict between different goods, all of which must be accepted, but which, taken together, cannot all be maximized. That is to say, we do not have a well-formed policy question, a fully formulated statement of a policy problem, until we are able to state the set of values or goods from which the question arises, and unless we are able to state that set of values or goods, so that we can discern their mutual inconsistency.

The issues involved in the contemporary movement for fiscal reform in education provide about as clear a model of policy questions generally as it is possible to shape. The policy issues are always "nested" within a set of mutually incompatible values or goods. We seek

(1) Equal educational opportunity for children.
(2) An equitable distribution of the tax burden.
(3) Local control of education.
(4) Responsible management of the State budget.

Maximizing any one of these goods--that is, getting as much of it as we can--will do damage to the advancement of the others. The policy problem is generated by the fact that we accept all four of these aims and yet they cannot all be maximized. We cannot have all the local control possible because doing so will probably mean getting less than would be good in the way of equity for children and taxpayers and control on the public budget. On the other hand, if we maximize equity for children, then we are likely to get more inequity in the tax burden and less local control. The problems of educational finance policy, in short, do not arise merely from the need to establish a more equitable system for taxpayers and children. They arise rather from the need to do so within a system of public goods that secures also both local control and responsible public management.

I daresay that all issues that we would describe as questions of public (or even personal) policy have this feature.[2] They are always "nested" in a set of social values or social goods which must all be considered, but which, taken together, are more or less mutually incompatible. Consider the issues surrounding the imposition of exit standards at the secondary school. Here we seek the mutual benefits of:

(1) Universal attainment (or at least racially and ethnically balanced frequencies of attainment).
(2) Common standards of achievement.
(3) Culturally pluralistic communities.

The mutual inconsistency of these aims is transparent when they are visibly juxtaposed. The difficulties of finding some balance in the entire set is well illustrated by recent experience in Florida and Virginia.[3] Policies are established to maximize common standards of achievement. The immediate consequence is to pay a price in securing the other two goods. So the policy is adjusted--or its implementation is delayed. It is not unreasonable to surmise that continuing adjustments will result in policies that appear on the surface to depart substantially from previous formulations, but whose consequences are not substantially different. Such a result is likely to emanate more from the presence of mutually incompatible values or goods than from administrative "bungling", or blindness, or from inefficiency, or from political chicanery. That is to say, neither the most efficient action nor the most technically profficient analysis will suffice to resolve the central conflict between the social aims within which the policy question resides. In general, there is no technical solution to a policy question. For example, there is no purely technical resolution of the fact that if, in social institutions, we get all the efficiency we can, then we are likely to have less community than we need or desire.

This conclusion, however, may seem outrageously facile. It deserves some explanation, and that explanation can be discovered in two points. The first requires that we grasp the important fact that what counts as an answer to a policy question always takes the form of a "What we should do" and never a "What we know". Only practical questions are admissible in a public forum, never theoretical questions. And this is fortunate indeed. It means that in the domain of policy we are able to arrive at agreement on what to do without having to agree on the reasons for doing it. We must be able to agree on a line of action and stick to it, even when we do not agree on what is good and even when we have different goals.[4]

The result of a policy question is always a decision and an action. The result of a theoretical question is always a truth claim. Policy deliberation is aimed at action, not at the acquisition of knowledge; theoretical questions are always aimed at the acquisition of knowledge, not at action. I do not mean by this claim that we can or ever should make public decisions without knowledge. Social action, no doubt, should be informed.

Nor do I mean that we can ever gain greater knowledge without some action. Research, after all, is a kind of action. I mean only that wise policy is never made with enough knowledge to *determine* a decision, and policy questions are never asked out of a primary interest in adding to our knowledge.[5]

In this day and age it is not hard to imagine someone saying, "If we just had a methodology sufficiently sophisticated and a body of relevant data sufficiently refined, then we could answer whatever policy questions may come along." Such a person has been captured by a delusion. The delusion consists in supposing that a policy question is a theoretical question when, in fact, it is not. Any time we suppose that a policy question can be *resolved* by some addition to our knowledge, then it will turn out that what we supposed was a question of policy has turned out to be merely a problem of engineering or efficient administration instead.

My point then is not that we should abandon all attempts to improve our methods of evaluation or policy analysis. My point is rather that since our indecision in matters of policy does not arise from the lack of such methods, therefore, it is unlikely to be laid to rest by their development. In matters of policy, we are confronted with indecision not because our knowledge or technical facility is faulty but precisely because we are confronted with a kind of question that, *in principle*, cannot be answered by any increment or improvement of knowledge. Our answers to policy questions may be improved by better information and better analyses in the sense that they will be more rationally persuasive. But such questions can, will, and usually are, answered even without such information. Furthermore, it is not obvious that the answers given in the absence of such analyses are worse than or even often different from the answers that would be given in their presence. In short, the answers may be "better grounded" rationally, but still not different or better in any other sense.[6] We can, no doubt, do something more rationally persuasive than consulting chicken entrails, but we are unlikely to get anything that in its results is quite as decisive. And this is so because of the properties of policy questions, not because of deficiencies in policy evaluation.

This observation brings me then to a second important reason why there are no technical solutions to policy problems. In paradise, there is no policy--except, perhaps, admissions policies. But why do policy questions not arise in paradise? There are many ideas of paradise, of course, but on the whole, one is inclined to view that men conceive of it as a perfected state of affairs in which wants presently denied in an imperfect world will somewhere or sometime be satisfied in a perfected world. Consider the view Emerson expressed in his essay "On Compensation". He heard a sermon, the chief message of which was that those saintly and good souls of the world who suffer without the comforts and amenities of life should persevere nonetheless in their goodness. For though sinners, with their fine carriages and furs, may seem to prosper now, they shall suffer later; and though saints may have to do without, they shall later be rewarded. The import of such a sermon, thought Emerson, was the message of saint to sinner, "You sin now. I shall sin later. I would sin now, but I can't." This is one rendering of the view that heaven is that perfected existence in which wants present, but now denied, will be satisfied.

All this suggests, of course, that the problem of optimality--therefore the need for policy--would be banished from any world in which human wants or desires are perfectly balanced by their satisfactions. What else could paradise be except a condition in which all human desires are satisfied?

There are two general strategies always sufficient to produce such a solution. The first lies on the side of doing something about the satisfactions available to human beings (productivity), and the second lies on the side of doing something about their desires (education and the development of character). The first leads us always to solve the problem of optimality by the provision of abundance; there is presumably no problem of satisfying wants. There is enough of everything, including enough justice and enough virtue. When there is no problem of

173

satisfying wants, there is no conflict of goods and, therefore, no problem of optimality.

The other strategy is the converse. There is no scarcity of what nobody wants. So the second way to resolve the problem of optimality lies not in the satisfaction of wants, but in their control, their composition. Thus, for Ghandi, diamonds and mink were plentiful, not because they were any the less scarce, but because they were not wanted. And not being wanted, they were abundant. If heaven is that condition in which wants are satisfied, then there may be abundance in heaven not because goods are maximized, but because wants are "composed". In neither case do problems of policy arise, and the reason they do not arise is that such methods of composing the goods of the world or of reconciling wants and satisfactions would render choice unnecessary. There is no conflict of goods in paradise. And the fact that there is consequently no need for policy is part of the proof that to formulate a policy question is to formulate the conflict of goods within which it is "nested".

But why do we not imagine paradise to contain interpersonal conflicts of wants? The answer is that to do so would involve either the judgment that wants are improperly controlled or that goods are insufficiently supplied to satisfy them. Such a condition would introduce a problem. Paradise would no longer be a perfected existence. It would contain the problem of optimality, a kind of allocational defect. Paradise would contain the problem of composing the most satisfactory combination of what goods do exist, and who should get them in what degree. In short, such a condition would introduce into paradise precisely those circumstances that create the need for policy and that dictate the features of any well-formed policy question--What are the goods in conflict? What is their best *possible* adjustment? How can we reach it? What are the trade-offs?[7]

The main point of this apparent digression is that problems of policy are an immediate and direct reflection of some immensely fundamental characteristics of the world and of human existence within it. They arise because the goods--not simply the interests--that human beings seek to secure in the world are interdependent and *do* conflict, not all the time in every respect, but all of the time in some respects. Only in paradise can we imagine that all human goods are simultaneously in sufficient supply so that there is no conflict in their allocation. It is important to note that since knowledge is a certain kind of good, then the idea of paradise includes the assumption that there is sufficiency of knowledge. But paradise does not arise *because* there is an abundance of knowledge. Policy questions are not banished from paradise because our capacity to know is perfected. They are banished rather because there is either an abundance of *all* goods or because there is a suitable composition of desires.

This formulation, however, is not all that is needed in exposing the presuppositions of policy questions. It deals with the presupposition of scarcity, but not with the presupposition of interdependence between goods. It is sufficient to show that when scarcity is absent--as it is in heaven--then no policy problems can arise. But the example presupposes that it makes conceptual sense to suppose that all human goods can exist in abundance *simultaneously*. Our imaginary hypothesis assumes that abundance of some goods can always be secured without significant costs in others.

The fact is, however, that human goods do conflict in such a way that they cannot all be provided simultaneously in sufficient supply to satisfy human desires. The point is central. Human goods do conflict so that the price of securing the abundance of some is always failure to secure as much as we would like of some other. Thus, if we succeed in providing as much equality as is wanted, we are unlikely to have as much liberty as is wanted. If persons develop as much tolerance for ambiguity as is wanted, we are unlikely to have as much courage as is wanted, and the price will sooner or later become apparent.

The view then is that human goods conflict (read "values" if you wish) not simply because they are in short supply, nor simply because interpersonal preferences conflict, but simply because they are not structurally consistent. They cannot all be maximized, even in paradise. It is a familiar idea that human wants or human interests conflict. But the view here is that human goods conflict. The fact that human interests conflict is what produces political problems, finding an adjustment between conflicting human interests. But the fact that human goods conflict is what produces policy questions, finding an adjustment between conflicting goods. Human goods continue to conflict even when human interests do not. Even when all are agreed on a single predominating interest--victory in an all-out war, for example--there will remain policy problems.

In other words, there are no unlimited goods in the world. There are no goods which, if provided in great abundance, would not have the consequence of certain other goods being in short supply. The ultimate solution to any problem of policy is, therefore, to be found only in paradise; that is, only under conditions in which problems of policy are not so much solved as they are simply non-existent. Such a state of affairs may be ideal; and in that respect, it may be optimal. But it is not possible, and in that respect, it is not optimal at all. That the ultimate solution of all policy problems is to be found only in paradise, may be precisely the fact that gives rise to the consistent and apparently inerradicable human impulse to think of social solutions to policy problems in utopian terms.

But utopian thinking is defective not merely because it pays too little attention to feasibility. It is flawed more fundamentally because it pays no attention at all to politics.[8] It is a fact of large significance that no well developed literary exposition of Utopia ever includes an account of politics. The central assumption always is that in Utopia the needed balance between conflicting human goods is resolved. Reconsideration is not needed. Therefore, the introduction of politics into Utopia would be a threat to--not a part of--the good life.[9] In Utopia, problems of policy remain only to the extent that there remain problems of monitoring the society and managing its affairs. Politics is replaced by administration, and the role of evaluation, if it exists at all, would be reduced to serving the ends of management.[10]

Policy questions do not arise at all in paradise. Serious ones do not arise in Utopia. But the reasons are different. The reason that serious ones do not arise in Utopia is not that goods are abundant or that desires are composed, but that the inherent conflict between goods is taken as resolved. All that remains is management. Not even the presumed Utopia of putting evaluators in charge would alter that result.

We must note a related feature of policy questions generally. Like a reporter filing his story for the evening edition, whoever will answer a policy question, in the real world, must do so within strict constraints of time. The reporter files a story by deadline, but always with the knowledge that there will be another deadline, and the present story can be amended by the next as events change and further facts are revealed. Two points are discernible in this observation. The first is that policy questions generally are answered in anticipation that the answer will be revised. The second point is that they are the kinds of questions that have to be answered on time, even though the information needed for the answer is not on time. Both points arise from temporal constraints, but they have different implications.

The first implies simply that policies are impermanent. We expect them to change. Often, they are not even very durable. They are not supposed to be. In that respect, policy questions are unlike constitutional questions, and they differ from moral questions in exactly the same respect. We do not expect persons to change their moral principles*--the constitution of their character--with great frequency. We do, however, expect policy to change with experience and with fair frequency. "Policy" implies "politics" and "polity"--a point that I shall consider in more detail later. We may note, for the moment, however, that if there is a practice whose improvement would promise the largest marginal gains in the formation of policy, it would be an improvement in the practice of politics, not evaluation.

The second point is equally vital. It means that just as it is better for the reporter to file a story on time without all the facts than to get all the facts and file the story too late, so also it is better, in the case of policy, to make a decision on time, but without all the facts, then it is to get all the facts and make the decision too late. In the case of policy, decisions have to be made always within large limits of uncertainty. Some reduction in the degree of uncertainty will be helpful, but the degree of reduction normally required for academic research is both improbable for policy decision and would often be undesirable even if it were not improbable.[11]

In other words, crude data arriving on time are always better than refined data arriving too late. So it is acceptable, even fortunate, that the methods required for policy decision are crude, even though the usual methods of research are necessarily refined. To answer a policy question, we need as much information as we can get. But "as much as we can get" usually turns out to be less than we could get if we had more time and, at the same time, more than can be used and more than will make a difference to the decision. Policy questions, in other words, are always answered in the midst of uncertainty, and there is always a point beyond which more information--however more excellent--will contribute little to the reduction of that uncertainty and do nothing to alter the direction of the decision.[12]

All this is simply another aspect of the claim that policy questions are practical rather than theoretical. They are questions of the sort that need to be answered, and that will be answered, even when we do not know which among alternative answers is the best. If these observations are credible, then it is possible to understand the claim sometimes advanced that

---

*Note the offense involved in calling them "policies"!

academic research is useful for policy decision inversely to its excellence as academic research.

But there is another, and more far-reaching, implication of these last observations. Suppose we define the character of a professional--as I think we must--in relation to the degree to which the professional's practice requires accuracy of approach in the midst of uncertainty. If we imagine a two-dimensional space defined by (x) increasing uncertainty in the predictable behavior of the materials dealt with in any practice, and (y) increasing uncertainty in the consequences of one's actions, then we would be able to array the practice of professionals, craftsmen, artists, and technicians as dispersed along the diagonal.

I would suggest (and this is only conjecture) that such a definition of "professional" would capture important features (though not all important features) of what we ordinarily mean by the term. But such a definition has interesting and important consequences. It means, among other things, that the education of professionals is an education in being able to make judgments and decisions in the context of large uncertainties. Such education, in a deep sense, is education not in technical skill (nor without technical skill), but in the capacity to deal with uncertainty, to live with doubt, to change one's mind. It is always ultimately a kind of education in self knowledge--learning one's limits.

But more importantly, and more concretely, it would tend to rearrange our conceptions of the relation between the practices of crafts and professions. For example, it would mean that the neuro-surgeon, although possessed of greater technical skill and manual dexterity, nevertheless, deals with a more predictable problem than the nurse, who, having to deal with the whole patient, acts in the midst of larger, though often less serious, uncertainties of both material and consequences. In short, by such a view, the nurse turns out to be engaged in a practice that is more professional than the surgeon's.

In an analogous way, the implication is that the evaluator, insofar as he presses for greater certainty and actually seeks to become the determiner of policy, is less of a professional, *at his best*, than the politician or executive, *at his best*. The drive of evaluators, to whatever extent they seek to find in practice a means of resolving policy questions, is in fact the drive to make the politician an evaluator, which is to say a technician of policy decision. Such an achievement, if ever realized, would constitute the most radical transformation of our political institutions and the practice of policy decision that one can imagine. Even if it could happen--which is doubtful--it would be undesirable if carried very far. But this discussion cannot be extended, refined, or made convincing without a further set of distinctions.

II

Facets of the Policy Process

Between policy analysis, policy formation, policy decision or promulgation, and the political analysis of policy there lie clear differences, and the practice of evaluation will relate differently to each. The tendency exists to regard these four activities--analysis, formation, decision, and political analysis--as steps in the policy process. But that view is misleading, because these activities are never fully discrete in practice and they do not occur in any persistent sequence. Nevertheless, there is a distinction of practice corresponding to each activity, and each practice, moreover, has its distinct kind of theory.

Policy Analysis

Policy analysis can be defined as the rational or technical assessment of the net marginal trade-offs between different policy choices. The question becomes "Which set of values will be advanced, which will not, and with what net benefits?" This is the same kind of question that we confront, say, in the design of a hand drill. What should be the design? The question is "nested" in a set of values. We want low cost, high safety, ease of handling, and durability. We can ask and rather precisely determine what marginal gains in one of these values will produce what corresponding costs in the others. If we "go for" greatest durability, then we are likely to get a higher cost and less ease of handling. If we "go for" the lowest cost possible, then we are likely to sacrifice something in the way of durability and safety. The design problem is to discover a balance between these competing values.

Enter the problem of incommensurabilities! How do we determine which among the competing values is to be given greatest weight? Which has the greatest worth--low cost, safety, durability, or ease of handling? Not even the most refined analysis of the costs and benefits will solve that problem. Such an analysis gives us the possibilities or a set of choices, but it does not pick out any preferred answer from within the set. Yet we need some procedure for doing just that. In short, we need a market decision, and getting a market decision is, no doubt, going to require a market analysis.

Is our market made up of professionals? Or does it consist essentially of amateurs and household craftsmen? If it is the former, then the problem will probably be resolved on the side of durability and safety with a slightly higher price. If the latter, then, by all means, the decision will probably be to minimize cost and sacrifice durability and, to some degree, safety. But then again, the market decision might be to "go for" the whole range of the market. Produce a variety of designs representing the full range of choices revealed by the analytic exercise. Something for everybody!

These activities are roughly analogous to the distinctions I want to make in the case of policy. Merely setting forth the marginal costs and benefits of a range of choices is one thing--policy analysis. Selecting one balanced choice or a range from within the possibilities is another thing--policy formation. The decision as to which choice or choices will be made is still a third--policy decision. And performing the market analysis needed for that decision is yet a fourth--policy analysis.

For example, suppose we entertain the prospect of distributing educational assistance to students in preference to institutions and that we are resolved to do so on the basis of financial need. In that case we require access to financial information, and not simply on groups, trends, or categories of persons, but on each actual individual. If we propose this kind of policy as more just than other choices, then, in the name of justice, individuals will have to reveal personal information that may have been regarded before as privileged. Two values conflict. We extend justice, but diminish, in some measure, privacy. To secure a definable gain in one, we pay a definable cost in the other. Policy analysis asks "What is the net marginal gain?" A truly refined policy analysis, which rarely exists, would tell us how much we are likely to gain in the advancement of justice for some corresponding cost in privacy. But no such analysis, no matter how refined, will tell us whether it is worth it. In order to resolve that question, we need something corresponding to a market analysis and a market decision. We need a political analysis and a political decision. "Policy" implies "polity" and "politics" just as "good industrial design" implies a structure for marketing analysis and marketing decision.

But consider another example. A Congressman asks whether pass-through requirements for allocating Title I (ESEA) funds should rest on tests of educational need rather than economic need. The answer comes back couched not in terms of "whether we should" but in terms of "what happens if we do?" That's policy analysis. In either case, the funds would go roughly to the same school districts--but not quite. What's the margin of "not quite"? Is "not quite" "very much"? Is it "enough to matter"? And even if it is "not much' would the change create incentives for local districts to pay more attention to "educational need" in answering allocational problems? And if so, then would the incentives be enough to make a difference? And if so, then (here we are again) how much of a difference? That's policy analysis.

But policy analysis does not, and need not stop there. It can ask not simply what the net consequences would be of doing X, but what those net consequences would be compared to doing Y, where Y is either what we are doing already or some third alternative.[13] The question for policy analysis is not whether doing X is a net improvement over doing Y--better than doing Y--but simply, what are the net effects? Whether it is better to have a drill of low cost instead of high durability will not be determined simply from an analysis of the trade-offs. It requires a marketing decision. Similarly, whether given the different consequences, it is better to do X than Y in public policy will not be determined by a policy analysis. It will be determined by a political decision resulting from a political process involving a political analysis.

In short, policy analysis is that rational, technical, analytic performance in which the central question is not whether X is a good thing to do, but simply what are the marginal effects of doing X, and what are the marginal effects as contrasted with doing something else instead? Hence, policy analysis is simply an activity whose theory is the theory of marginal utilities. It is, by all accounts, an activity that consists in the exercise of theoretical, rather than practical, rationality. It assumes that the policy question is "nested" in a conflict of values present as objective states of affairs in the society. It is an activity in which evaluators may take a leading role provided that they do not suppose they are actually evaluating policy, as opposed to merely recording--either in prospect or in retrospect--the consequences of doing X or Y.[14]

Policy Formation

Policy formation is an activity of a contrasting genre. Policy formation is that activity by which we seek to gain agreement on what form a specific policy can or will take, as opposed to what form it ought to take. Not even by the most refined policy analysis will we have actually formed a policy statement. Indeed, policy analysts are not typically in a position to actually formulate policy.[15] For the latter, we need to engage in conversation, persuasion, argument, and in (seemingly) endless meetings with those who will actually pen the regulation, mark up the bill, establish the procedures, write the guidelines, etc. The theory of policy formation can then be discerned as one aspect of the theory of government management and rhetoric. At the Federal level, it usually turns out to be the theory of inter-agency politics. "Don't fight over turf; just take up space" is a rule for the conduct of policy formation. I include here the theory of rhetoric because clearly it makes a difference what things are called. The same policy that under one name may never see the light of day will, under another name, pass without objection. Calling it "school aid" may defeat it; but calling the same thing "national defense" may insure its acceptance. "If it matters what you call it, then call it something that matters" is another guiding rule in the theory of policy formation.[16]

Policy Decision

Policy decision can be described as the authoritative action of some office, administrative or legislative, by which a line of action, for the moment at least, is established. Policy decision is not so much an activity or process as it is a momentary end point in the continuing business of government. It is that end point that is sometimes supposed by the naive to capture the entirety of the policy process--as though making policy could be reduced simply to an act of will or the result of divination. The theory of policy decision is simply the theory of the policy itself. It is the political and legal theory by which authority is distributed, obligations for decision are assigned throughout the structure of political institutions, and agents of authority are enjoined to act.

Political Analysis

Unlike policy analysis, political analysis is concerned not with determining the net benefits of a given course of action, but with their political weight.[17] The aim is not so much to determine the net social benefits of a particular policy, but to determine its constituency. If policy analysis is concerned with establishing what is good, then political analysis is concerned with estimating who will vote for it, whether the best thing to do is the same as the best thing that can be done. Hence, the theory of political analysis is the theory of political behavior.

We may gather these thoughts together in a brief culminating summary. The theory of policy analysis is the theory of marginal utilities. It establishes the set of policy choices.[18] The theory of policy formation is the theory of inter-agency politics. It is the governmental process by which a course of action comes to be selected. The theory of policy decision is nothing less than the theory of the policy itself, and the theory of political analysis is the theory of political behavior. When we view all of these activities together, not as discrete steps in the policy process, but as distinct facets of a social process--not one feature predominating and now another--then we can discern more clearly when the professional practices of evaluation fit and what their relevance is to the creation, promulgation and implementation of public policy.

Evaluators and evaluation can contribute to each of these activities, but not to each in the same way. For example, the rational standards of policy analysis are the standards of theoretical reason, but the rational standards of policy decision and political analysis are the standards of political judgment. These are practical activities. This difference may help to explain why it is that when the question, "What should we do?" is given a policy analysis, we may get one answer, and when given a political analysis or when rendered in a policy decision, we may get an entirely different answer. In short, the exercise of political judgment is a practical activity. It is also an evaluational activity. But the result of that activity may differ from or even contradict the results of policy analysis. What we should do--even the best thing to do--may turn out to be one thing by policy analysis and a very different thing when it comes to political decision.

The professional evaluator can contribute in the context of government, but he will contribute to all of these activities only to the extent that the evaluator becomes also a politician and a political advisor. Consider, for example, the case of policy formation. The evaluator, as professional, can contribute, but that contribution will be most substantial to whatever extent he becomes a student of bureaucracy and a trusted counselor to authoritative leadership.

So the dilemma is this. Each of these activities involves evaluation in some broad sense of the term. Each involves evaluation in the sense, say, that buying a camera does. But only in the case of policy analysis is the evaluator's role, as professional, undiluted by the need to take on other roles. The evaluator, as evaluator, is likely to make a contribution only to the conduct of policy analysis. But in government as elsewhere, the possession of knowledge can bring with it a certain kind of power. To the extent that the evaluator goes beyond his professional practice and with superior knowledge also earns the confidence of political leaders, exercises political judgment, and acquires the additional skills of a practiced political observer of the present bureaucracy and an uncertain future, then he will contribute to every facet of the policy process. But in doing so, he will also become less an evaluator in any limited professional sense and more a political leader or public servant in a quite old-fashioned and conventional sense. His main characteristic will not be the possession of technical skill. It will be the possession of civic virtue.

POLICY AND EVALUATION:
A CONCEPTUAL STUDY

Appendix

The six papers upon which this text was developed are:

1. Gordon Ascher. The Interaction Between Policy and Evaluation Within Oregon Department of Education.
2. Gerald W. Bracey. The Relationship Between Evaluation and Policy in the Virginia Department of Education, (Draft).
3. David L. Donovan and Stanley A. Rumbaugh. The Interaction Between Policy and Evaluation Within the Michigan Department of Education, (March 3, 1980 Draft).
4. James H. Gold. The Interaction of Policy and Evaluation in State Departments of Education, (March 1980).
5. Alfred Rasp, Jr.. Evaluation and Decision Making in Washington State.
6. Paul Sandifer. The Interaction Between Policy and Evaluation Within the South Carolina Department of Education, (October, 1979).

In the notes that follow, reference is made to these papers by the authors' last names. Numerical references are to points within the text of the conceptual paper by T. Green.

1. Bracey gives about as good a review of the confusion surrounding the term "policy" as I have seen. He, together with Gold and Sandifer, observes that policy is sometimes framed without "clear, procedural implications" and also that procedures are sometimes viewed as policy. Ascher tends to see policy as pretty much limited to authoritative rules of procedure, and nearly all of the authors, at one time or another, speak of policy as something akin to "personal policy of a person in authority".

   It has not seemed to me fruitful to aspire after too strict a definition of "policy". But to see what it means in the context of practice might be useful. Hence, I have tried to avoid what would appear to be a fruitless "academic" exercise of little practical significance by asking "What constitutes a policy question?" instead of "What is policy?"

2. Do all policy questions have these features? I think they do, but the papers do not clearly reveal that fact. On the contrary, it is pretty difficult to find any policy

questions really carefully formulated in these papers. What one finds, by implication, is a range of descriptions of technical and political problems surrounding some significant events in the history of State Department activities. The values in conflict are almost never fully drawn out. But they can be discerned inductively at work in the narratives provided.

For example, both Bracey and Gold remark, but in different ways, that when there is consensus, it can be decisive in resolving policy questions. But they also indicate that, in effect, this proviso, amounts to saying, "If all people agree on what is to be done, then it is no longer a policy problem" (Bracey, p. 30). The implication is that a conflict between values in which the policy issue resides is an essential feature without which such questions would not be serious.

Sometimes we observe from the narrative the points at which the "nested values" of the policy question begin to emerge through time. (See Bracey, p. 45, where they begin to emerge, but do not ever take the shape of a well-formed policy question. See also Bracey, p. 47. See also Donovan and Rumbaugh, pp. 55-56, where it becomes apparent that the policy issues arise as larger numbers of goods are permitted to enter in conflict through constitutional change and as their presence becomes more evident. Again, the same is evident in Donovan and Rumbaugh on p. 64, and pp. 72-73.)

This feature of policy questions generally is also expressed in Gold, p. 84. It is the essential requirement that he lists there as Contextual variable #3. The same can be seen in the South Carolina experience over legislation proposed in 1977. As Sandifer tells it, the policy question, to the extent that there was one, was nested in conflicting values on the part of legislative representatives.

3. The experience referred to here is the experience over the exit standards at the secondary level in Florida and the account given by Bracey on identifying the "cut-off score" for the initial tests in Virginia. In that latter case, consider the extent to which political considerations would, and probably should, enter quite beyond any considerations of technical decision, and the ways in which the necessity of those considerations reflect values in conflict and define the policy question. In short, as Bracey describes it, the policy question was not really whether there would be a cut-off score (or what it would be) but how to strike the appropriate balance between conflicting values at issue in that decision. The result was initially a decision without data. There followed a readjustment of that decision, and it ended up being very different in appearance, but not very different in consequences, I

suspect from what existed before. (See also Sandifer pp. 159-160.)

4. Is there any evidence in these papers of the assumption that people would agree on policy if they could only agree on goals? I am not sure. It is interesting, however, that those papers in which policy is seen to be most nearly associated with management, monitoring, and issues of compliance, (Donovan and Rumbaugh, Ascher, and Rasp) are also the papers within which issues of policy are most explicitly perceived as issues of management and administrative guidance.

5. This observation running through the paper is clearly recognized, though not this explicitly, in Sandifer in his extended comments on the problems of framing evaluations as though they were addressed to academics. (See Sandifer, pp. 155-156, and again p. 157.) But this point is also explicitly addressed in Bracey.

6. In Bracey's account of the program in Virginia, he provides clear instances in which policy decisions are made without data, with the suspicion that decisions might have been better with it, but with another suspicion that they might not be different in either case. Still, the view prevails both in Bracey and in Donovan and Rumbaugh that the possession of such data not only makes decisions more rationally persuasive, but may be necessary because of statutory requirements and for allocational decisions even when it does not produce a different result.

7. This notion that a well-formed policy question always contains these kinds of questions is not something that is well displayed in the papers. It would be interesting to take either a truly serious policy question (such as is implied in the personal goals of the SPI in Washington or issues of remediation as displayed in the behavior of the system in Virginia) or a matter of procedure as implied by policy decisions in other of these papers, and really examine what the policy question is, rather than, as is usually the case, consider merely the result of a policy decision and describe its operation.

8. Can one find in these papers the residue or evidence of utopian thinking? What would it look like? Well, for one thing there is in general a failure to take seriously in these papers the presence or the need for a political process. Where the presence is acknowledge, as in Bracey, Gold, and Sandifer there is also, it seems to me, a general failure to take that role seriously. Exception would be the Virginia paper. But Sandifer presents an interesting test of this. On p. 158, he explicitly acknowledges the necessity of taking the political process into account in

trying to understand the role of evaluation. But he regards such influences as "external" (to the Department, I assume) and not as essential. The same ambiguity is expressed on p. 154.

Bracey's paper presents another interesting, sensitive and sophisticated example of this kind of acknowledgement. But I wonder! would it be Bracey's opinion that things would be better on the whole if such "non-logical" behavior were not so influential? It would be the utopian impulse to say "Yes" and stop there. I suspect that Bracey (also Gold and Sandifer) would go on to answer that question with a "Yes, but . . . ."

9. I think it would be interesting to consider what features of evaluation within the operations of government regard the preservation of politics as something of an intrusion. Why is it not the view of evaluators that, after all, the play of political forces is the primary and only essential method of evaluation that we really have to preserve?

I believe that the answer to this question is that evaluators share a kind of utopian vision in which rational decisions replace political decisions. That is to say, the maintenance of politics is seen as an obstacle to the conduct of effective evaluation and an obstacle to making evaluation contributions to social decisions. In short, politics tend to be viewed as replaceable by management and management tends to be viewed as something that should be guided by evaluators.

10. The papers almost uniformly testify to the claim that the role of evaluation in state departments of education is to serve the ends of management and to keep politicians out of trouble. (See Rasp especially, on the last point.) This tends to be viewed by evaluation theorists as a defect, but it is usually viewed by those in state departments as their normal, natural, and rightful role.

11-12 All of the papers, with the possible exception of Ascher, comment on the stringent boundaries of time under which evaluators in state departments operate and the ways in which this fact marks a substantial contrast between what has to be done in the context of government and what can be done in the academic setting. Clearly, they are in agreement that this is one of the major differences that tends to make evaluation theory of little relevance to evaluation practice, at least as it occurs in State governments.

13. It is interesting, I think, that in none of these papers is there any detailed story about framing a "better than" kind of judgment in the case of policy analysis. Rasp remarks

that the section in Washington State never performs policy analysis. The same could be said of Ascher's account. But this is probably an expression of the fact that well-formed policy questions seldom surface in the context of State educational policy. That is, as I have observed already, the full explication of goods in conflict virtually never comes to the surface within state governmental affairs. If it did, as a matter of course, then there would have to be stories about "better than" kinds of judgments in the accounts given.

14. Policy analysis necessarily involves the study of marginal utilities and really excellent policy analysis (which seldom occurs) would require a study of net marginal utilities. One step in that kind of analysis will always be simply "finding out what happened"; in other words, it requires the study of effects. Note: "Study of effects" is what nearly all experimental and virtually all intervention designs are about. Effects can be "greater or less" more in one area than another, but the "study of effects", though an essential step, will never add up to a study of worth. GA gives a good illustration of this when he describes the circumstances in Oregon by which any decision of worth is "kicked downward" in the system. Studies of "effects", though essential as a step in policy analysis, remain nevertheless only a step. Evaluators in state departments contribute to the satisfaction of this step, but they cannot, by that means alone, complete a policy analysis.

15. None of the papers reflect the occupancy of a position sufficient to determine the actual formulation of policy.

16. The theory of policy formation could be described as the theory we use to predict the behavior of political and bureaucratic leaders. All of the papers, but especially Bracey and Rasp recount the significant difference that is created by the personalities of bureau leaders. Evaluation practice is substantially governed by the behavior of political leaders. So the theory of policy formation, as the theory of inter-agency politics, is likely to be expressed in the descriptions that political officials give and the descriptions that their subordinates give of their own personal qualities, and their own personal aims as actors in the political arena.

17. Notice that "political weight" is different from "rational weight".

18. The set of policy choices is also established by what are sometimes called "peremptory rules". (See Braybrooke and Lindbloom, A Strategy for Decision.) Such rules tend to establish the moral limits within which policy can be selected, but, at the same token, they tend to guarantee

that policies are to be chosen from among alternatives, all of which are morally permissible or have worth. In that sense, defining the set of policy choices is the expression of moral conviction and value estimations, but selecting from within that defined set is not. This is partly the reason why, with even the best evaluation data, we are unlikely to arrive at policy choices that are substantially different from those we would arrive at without such data. The range of choices is already substantially set by considerations that define the set of alternatives, and those considerations, being peremptory, do not permit a very large range of differences to arise.